

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 9, 2000, 20:47:27 ; Search time 1334.93 Seconds
(without alignments)
1927.840 Million cell updates/sec

Title: US-09-214-679-1
Perfect score: 1442
Sequence: 1 cccgggaactccatgtggcc.....aatgcaattcattggatcc 1442

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 972840 seqs, 892348106 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1945680

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:*

1: gb_ba1:*

2: gb_ba2:*

3: gb_cm:*

4: gb_ov:*

5: gb_pat:*

6: gb_ph:*

7: gb_pl1:*

8: gb_pl2:*

9: gb_pr1:*

10: gb_pr2:*

11: gb_pr3:*

12: gb_ro:*

13: gb_sts:*

14: gb_sy:*

15: gb_un:*

16: em_fun:*

17: em_hum1:*

18: em_hum2:*

19: em_in:*

20: em_om:*

21: em_or:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_sts:*

28: em_sy:*

29: em_un:*

30: em_vi:*

31: gb_htg1:*

32: gb_htg2:*

33: gb_in1:*

34: gb_in2:*

35: em_ba1:*

36: em_ba2:*

37: em_hum3:*

38: em_hum4:*

39: gb_pr4:*

40: gb_htg3:*

41: gb_htg4:*

42: gb_htg5:*

43: gb_htg6:*

44: gb_htg7:*

45: em_htg1:*

46: em_htg2:*

47: em_htg3:*

48: em_hum5:*

49: gb_pl3:*

50: gb_pr5:*

51: gb_htg8:*

52: gb_htg9:*

53: gb_htg10:*

54: gb_htg11:*

55: gb_htg12:*

56: gb_htg13:*

57: gb_htg14:*

58: gb_in3:*

59: gb_htg15:*

60: gb_htg16:*

61: gb_htg17:*

62: em_htg4:*

63: em_htg5:*

64: em_htg6:*

65: em_htg7:*

66: em_hum6:*

67: gb_htg18:*

68: gb_htg19:*

69: gb_htg20:*

70: gb_htg21:*

71: gb_htg22:*

72: gb_htg23:*

73: gb_htg24:*

74: gb_htg25:*

75: gb_htg26:*

76: gb_htg27:*

77: gb_htg28:*

78: gb_htg29:*

79: gb_htg30:*

80: gb_htg31:*

81: gb_vil:*

82: gb_vil2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1442	100.0	1442	5	A69475	A69475 Sequence 1
2	1442	100.0	1442	5	A72152	A72152 Sequence 1
C 3	20	1.4	1100	77	AC057459	AC057459 Giardia i
C 4	20	1.4	3167	2	RMU83439	U83439 Rickettsia
C 5	20	1.4	3176	2	AF018076	AF018076 Rickettsi
C 6	20	1.4	3176	2	RSU83455	U83455 Rickettsia
C 7	20	1.4	3176	2	TTU83456	U83456 Thai tick t
C 8	20	1.4	3179	2	RMU83447	U83447 Rickettsia
C 9	20	1.4	3182	2	AF120018S2	AF120019 Rickettsi
C 10	20	1.4	3182	2	AF120020S1	AF120020 Rickettsi
C 11	20	1.4	3182	2	AF120022S2	AF120023 Rickettsi
C 12	20	1.4	3182	2	RAU83436	U83436 Rickettsia
C 13	20	1.4	3182	2	RCU83440	U83440 Rickettsia
C 14	20	1.4	3182	2	RCU83443	U83443 Rickettsia
C 15	20	1.4	3182	2	RCU83448	U83448 Rickettsia
C 16	20	1.4	3182	2	RCU83453	U83453 Rickettsia
C 17	20	1.4	3182	2	RU83442	U83442 Rickettsia
C 18	20	1.4	3182	2	RP83449	U83449 Rickettsia
C 19	20	1.4	3182	2	RSU83452	U83452 Rickettsia
C 20	20	1.4	3182	2	RSU83454	U83454 Rickettsia
C 21	20	1.4	3185	2	ARU83437	U83437 Astrakhan r
C 22	20	1.4	3185	2	ITU83441	U83441 Israeli tic
C 23	20	1.4	3185	2	RAU83446	U83446 Rickettsia
C 24	20	1.4	3188	2	RMU83444	U83444 Rickettsia


```
Db 1141 ATACACCGTGGCGGATGCTGAACAAAACCTGTAGTTAGTAGGAATAACTAACCGG 1200
QY 1201 tgaacattaccggatgtagatcggggtaagtgtgaagtcaacaatacgctatttttaa 1260
Db 1201 TGAACATTACCGGATGATAGTCGGGTAATGTGAAGTTCAACAAATCGCTATTTTTAA 1260
QY 1261 cagctaaagcaggtgcatatggggccagatacaccatacaattatggtttactactcc 1320
Db 1261 CAGCTAAAGCAGGTGCATATGGGGCCAGATACACCATCAATATTGGTTTACTCTCC 1320
QY 1321 ttacagcaggtgacgcggcacaagagttgtcacaaatgagcgcgagcaaccagactatt 1380
Db 1321 TTCAGCGAGTGACCGGCCACAGAGTTGTCAATGGCGGAGCAACCCAGGCTATT 1380
QY 1381 gccgaattaatcaaaatggcgcatcaacgagcagaccactcaatgcaattcattggat 1440
Db 1381 GCCGAATTAATCAAAATGGCGGCATCAACGCGCAGACCACTCAATGCAATTCTTGGAT 1440
QY 1441 cc 1442
Db 1441 CC 1442

RESULT 2
A72152 1442 bp DNA PAT 11-MAY-1999
LOCUS Sequence 1 from Patent WO9801568.
ACCESSION A72152
VERSION A72152.1 GI:4808107
KEYWORDS
SOURCE Klebsiella oxytoca.
ORGANISM Klebsiella oxytoca
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Klebsiella.
1 (bases 1 to 1442)
Brieden,W., Naughton,A., Robins,K., Shaw,N., Tinschert,A. and
Zimmermann,T.
METHOD OF PREPARING (S) - OR (R) -3,3,3-TRIFLUORO-2-HYDROXY-2-
METHYLPROPIONIC ACID
Patent: WO 9801568-A 15-JAN-1998;
LONZA AG (CH)
FEATURES
source
1..1442
Location/Qualifiers
/organism="Klebsiella oxytoca"
/strain="PRs1"
/isolate="PRs1"
/db_xref="taxon:571"
197..1183
/codon_start=1
/transl_table=11
/product="AMIDASE"
/protein_id="CAB42722.1"
/db_xref="GI:4808108"
/translacion="MKWLEESIMAKRGVGRKPVTHLTEMKEFHVTIGPVSTPV
LTIIEPGRIVDTRDAFEGAINSEQDIPSOILKMPFLNPQNGIMVNGAEKGDVLAVY
IESMLPRVDYIGICAMIEPHFGTLTDLTAMLANDPLPEKVRMIKIDSEKGVWSKRHT
LPYKPHIGTSLVSPEIDSINSLTDPNHHGNMDVDPDIPGSIITYLPVPRAPEGRLFTGDA
HACODGEICSTAVEFASITTKVDLIKHWOLSWPMENAINISIGSRLEDA TRI
AYRLIYHLVDFEFGQWDAYMLLSQCKVRLGNVDPKYTVGAMLNKLIL"
BASE COUNT 385 a 350 c 370 g 337 t
ORIGIN

Query Match 100.0%; Score 1442; DB 5; Length 1442;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1442; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cccgggaactccatggtggcggtgactcctgtgcgagcagagatattgcgatgccagcg 60
Db 1 CCCGGAACTCCATGTGGCGGTGATCTGCTGCGACGAGATATTGGCATTCACCGCGG 60
QY 61 ccgacagcgtgtgcggttaattgataaaggcctggttagaagaactgacccaacaac 120
Db 1141 ATACACCGTGGCGGATGCTGAACAAAACCTGTAGTTAGTAGGAATAACTAACCGG 1200
```

```
Qy 1201 tgaacattaccggatgtagatcggggtaagtgttaagtccaacaatcgctatttttaa 1260
|||||
Db 1201 TGAACATTACCGGATGTAGATCGGGGTAAATGTGTAAAGTTCACAAATCGCTATTTTAA 1260

Qy 1261 cagctaaagcgggtgcataatggtggccagatcacaccatcaatattggttacttactcc 1320
|||||
Db 1261 CAGCTAAAGCAGGTGCATATGCGGCCAGATACACCCATCAATATTGGTTTACTTTACTTCC 1320

Qy 1321 ttcagcgagtgacgagcgcagagtgatgacaaatggcgcgagcaaacaccagcattt 1380
|||||
Db 1321 TTCAGCGAGTGACGCGCGCACAGAGTGTGCACAAATGCGCGGAGCAACCCAGCGCTATT 1380

Qy 1381 gccgaataataatgagcggatgacacgcagacacacacacacacacacacacacacacac 1440
|||||
Db 1381 GCCGAATTAATCAAAATGGCGCATCAACGCGAGACACCACTCAATGCAATTCATTGGAT 1440

Qy 1441 cc 1442
||
Db 1441 CC 1442

RESULT 3
AC057459/c
LOCUS AC057459 1100 bp DNA HTG 18-APR-2000
DEFINITION Giardia intestinalis clone M2849 strain WB-C6, LOW-PASS SEQUENCE
SAMPLING.
ACCESSION AC057459
VERSION AC057459.1 GI:7594157
KEYWORDS HTG; HTGS_PHASE0.
SOURCE Giardia intestinalis.
ORGANISM Giardia intestinalis.
REFERENCE 1 (bases 1 to 1100)
AUTHORS Morrison,H.G., McArthur,A.G., Nixon,J., Eakin,N.O., Kim,U.,
Crocker,M.C., Hinkle,G., Holder,M.E. and Sogin,M.L.
TITLE Giardia: a model for ancient eukaryotic genome analysis
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1100)
AUTHORS Morrison,H.G., McArthur,A.G., Nixon,J., Eakin,N.O., Kim,U.,
Crocker,M.C., Hinkle,G., Holder,M.E. and Sogin,M.L.
TITLE Direct Submission
JOURNAL Submitted (18-APR-2000) Josephine Bay Paul Center for Comparative
Molecular Biology and Evolution, Marine Biological Laboratory, 7
MBL Street, Woods Hole, MA 02543-1015, USA
COMMENT * NOTE: This record contains 1 individual
* sequencing reads that have not been assembled into
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1 1100: contig of 1100 bp in length.
FEATURES
source
Location/Qualifiers
1..1100
/organism="Giardia intestinalis"
/strain="WB-C6"
/db_xref="taxon:5741"
/clones="M2849"
BASE COUNT 268 a 299 c 222 g 311 t
ORIGIN
Query Match 1.4%; Score 20; DB 77; Length 1100;
Best Local Similarity 100.0%; Pred. No. 6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 225 ccaaacgcggtgtgtggtgcc 244
|||||
```

```
Db 229 CCAAACGCGGTGTGTGTGCC 210

RESULT 4
RMU83439/c
LOCUS RMU83439 3167 bp DNA BCT 10-SEP-1998
DEFINITION Rickettsia mongolotimonae rOmpA (ompA) gene, partial cds.
ACCESSION U83439
VERSION U83439.1 GI:1778884
KEYWORDS Rickettsia mongolotimonae.
SOURCE Rickettsia mongolotimonae.
ORGANISM Rickettsia mongolotimonae
Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
Rickettsiaceae; Rickettsiaceae; Rickettsia.
REFERENCE 1 (bases 1 to 3167)
AUTHORS Fournier,P.E., Roux,V. and Raoult,D.
TITLE Phylogenetic analysis of spotted fever group rickettsiae by study
of the outer surface protein rOmpA
JOURNAL Int. J. Syst. Bacteriol. 48 pt 3, 839-849 (1998)
MEDLINE 98404549
REFERENCE 2 (bases 1 to 3167)
AUTHORS Raoult,D., Fournier,P.-E. and Roux,V.
TITLE Direct Submission
JOURNAL Submitted (31-DEC-1996) Faculte de Medecine, Unite des Rickettsies,
CNRS EP J0054, 27 bd Jean Moulin, Cedex 5, Marseille, France
FEATURES
source
Location/Qualifiers
1..3167
/organism="Rickettsia mongolotimonae"
/strain="HA-91"
/db_xref="taxon:45261"
1..3167
/gene="ompA"
<1..>3167
/gene="ompA"
/codon_start=3
/transl_table=11
/product="rOmpA"
/db_xref="GI:1778885"
/db_xref="GI:1778885"
/translation="ITLHAGGSLDANNIDFGARSTLEFNGGGIPYFKAIANGNNA
ILNVNKKLTAYHLTIGTVAEINIGAGNIFAIDASAGDVTILNADIDHFRALDSTLLL
SNLTGCVNNILLAAADLVAPGDEGTVFNGGVNGLNIGSNVAGTARNIGGGGNKFN
ALLIYNAVITDDVLEIGQNLINNNADFTSSTAFNAGAIQIONATYITDANNGLIN
IPAGNTQFAHADAOQLILONSSGNDRTITLGANDIDPDNDDEGIVILNSVTAGKLTLAG
GKTFGAHKLQITVFKGAGDCSTAGTTFTNTIVLDITGOELGATVANYVLFKDAVG
LFTQGTIGTGFIDFNAGKVTLLNNNVNAGTVQNTGNTNGTLVLGASNLNVNGIA
MLKVGAGNVTIAKGNVKIGEIOGTNTLTLPAHNLTGSIINKTGOALKLNFMSG
SVSGVGTAAANSVDITTAGATSFASVNAKGTAKLGTTSPAHFTTNTGAVTLAKGS
ITSFARNVTATSEVANSATINFGNSLAFNSNITGSGTTLTLGANQVITYTGTGTDPL
TLNNTFDGAAGSGNLIKSGSTLDLSGVSTLALVVTATNFDNNISPDPTKYTVIAE
TVGGLKPTKENVKITINDNRFVDFTFDASTLTFLAKDIAADVIDKDFAPGGLANI
PNAANIKKSLKLMEDAPNGDARQAFNFGMLTPEOFADATTHLMODVVKPSTIAAV
NNOVASNTSSNITALNARMKVAGNKGPVSSGDEMDAKFGAWISPFVGNATQKMC
NSISGYKSDTGTGTIGFDGFDVDDLVLGLAYTRADDIDKLNNKTGDKKNVSNISKL
YGLISVPIENLFEATIASYSDNKSRRRVIAITLETVGTFANGKRYKSESTYGQLM
AGTYMMPENINLTPLVGURYSITKDKGYKETGTTQNTLVKGNKNTYFDGLLGAKVS
SNINVEILVTPELYAMVDFAKNKVSAIDARLQGMTAPLPTNSFKQSKTSFDVGVG
TAKYKMEYGINVDITNIGSKYFAQ"
BASE COUNT 1062 a 513 c 666 g 926 t
ORIGIN
Query Match 1.4%; Score 20; DB 2; Length 3167;
Best Local Similarity 100.0%; Pred. No. 6.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1373 aggcattatgcgaataat 1392
|||||
Db 1603 AGGCTATTGCGGAATAAT 1584

RESULT 5
AF018076/c
```



```
LOCUS AF018076 3176 bp DNA BCT 08-DEC-1998
DEFINITION Rickettsia honei rOmpA (OmpA) gene, partial cds.
ACCESSION AF018076
VERSION AF018076.1 GI:3982570
SOURCE Rickettsia honei.
ORGANISM Rickettsia honei.
REFERENCE 1 (bases 1 to 3176)
AUTHORS Stenos,J., Roux,V., Walker,D. and Raoult,D.
TITLE Island spotted fever in Australia
JOURNAL Rickettsia honei sp. nov., the aetiological agent of Flinders
MEDLINE Int. J. Syst. Bacteriol. 48, 1399-1404 (1998)
REFERENCE 99045882
AUTHORS Roux,V., Thiers,B. and Raoult,D.
TITLE Direct Submission
JOURNAL Submitted (11-AUG-1997) Faculte de Medecine, Unite des Rickettsies,
CNRS UPRES-A 6020, 27, bd Jean Moulin, Marseille 13385, France
FEATURES
source
1. .3176
/organism="Rickettsia honei"
/strain="RB"
/db_xref="taxon:37816"
<1. .>3176
/gene="ompA"
<1. .>3176
/gene="ompA"
/codon_start=3
/transl_table=11
/product="rOmpA"
/db_xref="GI:3982571"
/translation="ITLQAGSLAANNIDFGAGSTLDENGLPGGNAIPYYFKGAIAN
GNNAILNNTKLLTAYHSTIGTVAEINIGAGNLEIDRAGSDVTLINQAQDINFRALDST
LVLSNLTGVGNLILAAADLVAPGAGDEGTVFNGVNGNLIGSNVAGTARNIGDGGN
KNTPLIYNAVITDDVNLEGIQNLVINDNADFTSSTAFNAGTIOIKDATYITDANNG
FNLNIPAGNIOFAHADAOLILQNSGNDRTITLGANIDPDNDDEGIVILNSVTAGKILT
IAGKMGFGAAHKLOTIVFKGAGCGCTAGTFTNTNIVLDITGOLELGATPAAVLYFND
AVOLTGTGNIIGGFLDNKAGTNTLANNVAVCTVONTGNTNGTTLVLGASNLNRVN
GIAMKVGAGNVTIAGKGNVKGIEIOGTGNTITLPAHNLTSINTGGOALKLNF
NGSVSGVGTAAVNSVDITAGTATFSASNAGTATLGTTSTFANTFTNTGATVLA
KGSITFVKNVATSFVANSATINFGNSLAFNSNITGSGTTLTLAGNOVVTGTGSGFT
DITLNTFDGAAGSGNLIKSGSTLDLSCVSTLALVVTATNFDNMNISPDKYTVI
SATAGLKLTPPENVKITINNDNRFVDFCDASTLTPAEDIAADVIDEDFAPGGPL
ANIPNAANTIKKSLMEDAPNGSDARQANNFGLMTPLOEADATTHLMQDVVKPSDTI
AAYNQVVASNISNITALNARMDKQVAGNKGVPSSGDEMDAKFGAWISPFVSNATQ
KMCNSISGYKSDTGGTIGDFGVDDLALGLAYTRADTDIKLKNKTKDKNKVESNI
YSYGLYSPVENLFEVIAIASYSDNKIRSKSRVIATLTETGVGTANGKYKSESFTG
QLMAGYTYMMPENILNPLAGLRYSTIKDKGYKETGTYQNLTVKGNYNFDCGLIGA
KVSSNINVNEILVTPELYAMVDYAFKNKVSADARLOGMTAPLPTNSFKOSKTSFVVG
VGVTAKYMKMEYGINDTNIGSKYFAQQ"
BASE COUNT 1063 a 523 c 677 g 913 t
ORIGIN

Query Match 1.4%; Score 20; DB 2; Length 3176;
Best Local Similarity 100.0%; Pred. No. 6.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1373 aggcattgccgaaataat 1392
|||||
Db 1612 AGGCTATTGCCGAATTAAT 1593

RESULT 6
RSU83455/c
LOCUS RSU83455 3176 bp DNA BCT 10-SEP-1998
DEFINITION Rickettsia sibirica rOmpA (ompA) gene, partial cds.
ACCESSION U83455
VERSION U83455.1 GI:1778904
KEYWORDS
```

```
LOCUS AF018076 3176 bp DNA BCT 08-DEC-1998
DEFINITION Rickettsia sibirica
Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
Rickettsiaceae; Rickettsieae; Rickettsia.
REFERENCE 1 (bases 1 to 3176)
AUTHORS Fournier,P.E., Roux,V. and Raoult,D.
TITLE Phylogenetic analysis of spotted fever group rickettsiae by study
of the outer surface protein rOmpA
JOURNAL Int. J. Syst. Bacteriol. 48 pt 3, 839-849 (1998)
REFERENCE 98404549
AUTHORS Raoult,D., Fournier,P.-E. and Roux,V.
TITLE Direct Submission
JOURNAL Submitted (31-DEC-1996) Faculte de Medecine, Unite des Rickettsies,
CNRS EP J0054, 27 bd Jean Moulin, Cedex 5, Marseille, France
FEATURES
source
1. .3176
/organism="Rickettsia sibirica"
/strain="246"
/db_xref="taxon:35793"
1. .3176
/gene="ompA"
<1. .>3176
/gene="ompA"
/codon_start=3
/transl_table=11
/product="rOmpA"
/protein_id="AAC35191.1"
/db_xref="GI:1778905"
/translation="ITLHAGSLDANNIDFGARSTLEFNGLDGGIPYFKGAIAANG
NNAILNVTKLLTAYHLITGTVAEINIGAGNIFAIDASAGDVTLINQAQDIFRALDST
LVLSNLTGVGNLILAAADLVAPGAGGTVFNGVNGNLHIGSNVAGTARNIGDGGN
KFNALLIYNAVITDDVNLEGIQNLVINDNADFTSSTAFNAGTIOIQNATYITDANNG
NLANIPAGNIOFAHADAOLILQNSGNDRTITLGANIDPDNDDEGIVILNSVTAGKILT
IAGKMGFGAAHKLOTIVFKGAGCGCTAGTFTNTNIVLDITGOLELGATPAAVLYFND
AVOLTGTGNIIGGFLDNKAGTNTLANNVAVCTVONTGNTNGTTLVLGASNLNRVN
GIAMKVGAGNVTIAGKGNVKGIEIOGTGNTITLPAHNLTSINTGGOALKLNF
NGSVSGVGTAAVNSVDITAGTATFSASNAGTATLGTTSTFANTFTNTGATVLA
KGSITFVKNVATSFVANSATINFGNSLAFNSNITGSGTTLTLAGNOVVTGTGSGFT
DITLNTFDGAAGSGNLIKSGSTLDLSCVSTLALVVTATNFDNMNISPDKYTVI
SAAEAGLKLTPPENVKITINNDNRFVDFCDASTLILFAKDIADVIDKDFAPGGPL
ANIPNAANTIKKSLMEDAPNGSDARQANNFGLMTPLOEADATTHLMQDVVKPSDTI
AAYNQVVASNISNITALNARMDKQVAGNKGVPSSGDEMDAKFGAWISPFVSNATQ
KMCNSISGYKSDTGGTIGDFGVDDLALGLAYTRADTDIKLKNKTKDKNKVESNI
YSYGLYSPVENLFEVIAIASYSDIRKSKSRVIATLTETGVGTANGKYKSESFTG
QLMAGYTYMMPENILNPLAGLRYSTIKDKGYKETGTYQNLTVKGNYNFDCGLIGA
KVSSNINVNEILVTPELYAMVDYAFKNKVSADARLOGMTAPLPTNSFKOSKTSFVVG
VGVTAKYMKMEYGINDTNIGSKYFAQQ"
BASE COUNT 1056 a 517 c 671 g 932 t
ORIGIN

Query Match 1.4%; Score 20; DB 2; Length 3176;
Best Local Similarity 100.0%; Pred. No. 6.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1373 aggcattgccgaaataat 1392
|||||
Db 1612 AGGCTATTGCCGAATTAAT 1593

RESULT 7
TTU83456/c
LOCUS TTU83456 3176 bp DNA BCT 10-SEP-1998
DEFINITION Thai tick typhus rickettsia rOmpA (ompA) gene, partial cds.
ACCESSION U83456
VERSION U83456.1 GI:1778906
KEYWORDS
SOURCE Thai tick typhus rickettsia.
ORGANISM Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
Rickettsiaceae; Rickettsieae; Rickettsia.
REFERENCE 1 (bases 1 to 3176)
```

AUTHORS Fournier,P.E., Roux,V. and Raoult,D.
 TITLE Phylogenetic analysis of spotted fever group rickettsiae by study of the outer surface protein rOmpA
 JOURNAL Int. J. Syst. Bacteriol. 48 Pt 3, 839-849 (1998)
 MEDLINE 98404549

REFERENCE 2 (bases 1 to 3176)
 AUTHORS Raoult,D., Fournier,P.-E. and Roux,V.
 TITLE Direct Submission
 JOURNAL Submitted (31-DEC-1996) Faculte de Medecine, Unite des Rickettsies, CNRS EP J0054, 27 bd Jean Moulin, Cedex 5, Marseille, France

FEATURES Location/Qualifiers

source 1..3176
 /organism="thai tick typhus rickettsia"
 /strain="TT-118"
 /db_xref="taxon:45259"
 gene 1..3176
 /gene="ompA"
 CDS <1..>3176
 /codon_start=3
 /transl_table=11
 /product="rOmpA"
 /protein_id="AAC35192.1"
 /db_xref="GI:1778907"

/translation="ITLQAGGSLAANNIDFGAGSTLDFNGLGGGNAIPYFKGAIAN
 GNAAILNVTKLTAYHSTIGTVAEINIGAGNLLIEDASAGDVTILNAQDINFRALDSA
 LVLNLTGVGVKNILLAAADLVAPGADEGTWFGVNGVNLIGSNVAGTARNIGDGGGN
 KNTLSIYNVAVITDDVNLEIQNVLINADNADFTSTAFNAGTIOIKDATYTIIDANG
 NLNIPAGNTOFAHADQAQLLQNSGNDRIITLIGANIDPDNDEGIVILNSVTAGKKLT
 IAGKRFGGAHKLQTLIVFKAGNCGAGTFTNTNIVLDITGOLELGATVAVLVFN
 AVOLQTGIGGFLDNKNGVTNNVNVNAGTVQNTGGTNGGTLILVIGASNLNRVN
 GLAMLVKAGNVTIAGGNVKTIGTGTNTLTPAHFLNGTSINKTGGQALKLNF
 NGSVGVVGTAAANSVDITTAGATSFASVNAKGTATLGGTTSFANTTNTGAVT
 KGSITFAKNTATSPVANSATINFGNSLAFNSNITGSGTTLTGANVYTYGTGSGFT
 DTLTLTFDGAAGKSGNLLIKSGTLDLSGVSTLALVVTATNFDMNNISPDTKYVI
 SAEATGLATPKENVKITINDNRFVDTFADSTLTTLFAEDIAADVDEDPAGGPL
 AALPNAANIKKLEMDAPNGSDARQAFNFGMLTPQEAADATTHLMDQVVKPSDTI
 KAVNNQVANSISNTALNARMKVQAGNKGVSSEDMDAKFGAMTSPFVSNAIQ
 KWNCSISGKSTGTIGTGFVSDLLALGLAYTRADTDIKLNNKTGDKKVESNI
 YSVGLNYPYENLFEAJASDNKIRSKSRVVIATLTETGYOTFANGKYKESVTG
 OLMAGVYMPENINLTPLAGLRYSTIKDKGKETGTYONTLVKGNYNPDGLLGA
 KYSSNINVNEIVLPELYAMVDYAFANKKVSADARLOGMTAPLPTNSFKQSKTSFVVG
 VGVTAHKHMEYGINDTNIGSKYFAQQ"

BASE COUNT 1063 a 523 c 676 g 914 t

Query Match 1.4%; Score 20; DB 2; Length 3176;
 Best Local Similarity 100.0%; Pred. No. 6.6;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1373 agcgtattgccgaataat 1392
 |||||
 Db 1612 AGCCTATGCCGAATAAT 1593

RESULT 8
 RMU83447/c
 LOCUS RMU83447 3179 bp DNA BCT 10-SEP-1998
 DEFINITION Rickettsia montana rOmpA (ompA) gene, partial cds.
 ACCESSION U83447
 VERSION U83447.1 GI:1778890
 KEYWORDS
 SOURCE Rickettsia montanensis.
 ORGANISM Rickettsia montanensis.
 Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 Rickettsiaceae; Rickettsiae; Rickettsia.

REFERENCE 1 (bases 1 to 3179)
 AUTHORS Fournier,P.E., Roux,V. and Raoult,D.
 TITLE Phylogenetic analysis of spotted fever group rickettsiae by study of the outer surface protein rOmpA
 JOURNAL Int. J. Syst. Bacteriol. 48 Pt 3, 839-849 (1998)
 MEDLINE 98404549

REFERENCE 2 (bases 1 to 3179)
 AUTHORS Raoult,D., Fournier,P.-E. and Roux,V.
 TITLE Direct Submission
 JOURNAL Submitted (31-DEC-1996) Faculte de Medecine, Unite des Rickettsies, CNRS EP J0054, 27 bd Jean Moulin, Cedex 5, Marseille, France

FEATURES Location/Qualifiers

source 1..3179
 /organism="Rickettsia montanensis"
 /strain="M5/6"
 /db_xref="taxon:33991"
 gene 1..3179
 /gene="ompA"
 CDS <1..>3179
 /codon_start=3
 /transl_table=11
 /product="rOmpA"
 /protein_id="AAC35183.1"
 /db_xref="GI:1778891"

/translation="ITLQAGGSLAANNIDFGVSTLEFNGPLDGGDIIHYHFKGAIA
 NGNATLVNVTKSLTAYHSTIGTVAEINIGADSFFTIDASGDTVILNAQDINFRACN
 STLMLNLTGVGVKNILLAAADLVAPGADEGCYVFNGLNIGSNVAGTARNIGDGG
 GKFNNLFYINVVKYTDVYNLEGIKNVFIGNDADFTSSACNAGTQINNAYIADAN
 NGNLNVPAGNIQFVHAGAOVLVLONSSENDRTIIGANIDPDNDGDGIVILNSVTAGKKL
 TIAGGKTFGGAHKLQAIIVFKAGNCGAGTFTNATDIVLDTISQELGATTANVLLN
 DAVLQTQDIDIGFLDNARNGVTNNVNVNAGTVQNTGGTNGGTLILVIGASNLNSV
 NGIAMLKAGNVTIAGGVDITTAGATSFASVNAKGTATLGGTTSFADTFTWTGAVTL
 TNGSVGVVGTAAANSVDITTAGATSFASVNAKGTATLGGTTSFADTFTWTGAVTL
 AKGNTFAKNTATSPVANSATINFGNSLAFNSNITGSGTTLTGANVYTYGTGSGFT
 TDTLTLTFDGAAGKSGNLLIKSGTLDLSGVSTLALVVTATNFDMNNISPDTKYVI
 ISAEATGLATPKENVKITINDNRFVDTFADSTLTTLFAEDIAADVDEDPAGGPL
 LAEPNAAANIKKLEMDAPNGSDARQAFNFGMLTPQEAADATTHLMDQVVKPSDT
 IAAVNNQVANSISNTALNARMDEVAQNKGPVSSGDEMDKFGAMTSPFVSNAIQ
 OKMNNISGVKSDTGTGTFGGLSDLLALGLAYTRADTDIKLNNKMDKKNVESN
 IYSLGLYSVPYENLFEAJASDNKIRSKSRVVIATLETGYOTFANGKYKESGYT
 GQMGAGTYMPENINLTPLAGLRYSTIKDKGKETGTYONTLVKGNYNPDGLLGA
 ARVSSNINVNEIVLPELYAMVDYAFANKKVSADARLOGMAAPLPTNSFKQSKTSFDI
 GVGVTAKHMEYGINDTNIGSKYFAQQ"

BASE COUNT 1055 a 516 c 682 g 926 t

Query Match 1.4%; Score 20; DB 2; Length 3179;
 Best Local Similarity 100.0%; Pred. No. 6.6;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1373 agcgtattgccgaataat 1392
 |||||
 Db 1615 AGCCTATGCCGAATAAT 1596

RESULT 9
 AF120018S2/c
 LOCUS AF120018S2 3182 bp DNA BCT 16-FEB-2000
 DEFINITION Rickettsia sp. Dns28 OmpA (ompA) gene, partial cds.
 ACCESSION AF120019
 VERSION AF120019.1 GI:6979178
 KEYWORDS
 SEGMENT 2 of 2
 SOURCE Rickettsia sp. Dns28.
 ORGANISM Rickettsia sp. Dns28.
 Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 Rickettsiaceae; Rickettsiae; Rickettsia.

REFERENCE 1 (bases 1 to 3182)
 AUTHORS Rydkina,E., Roux,V., Fetisova,N., Rudakov,N., Gafarova,M., Tarasevich,I. and Raoult,D.
 TITLE Detection of New Spotted Fever Group Rickettsiae in Ticks Collected in the Former USSR
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 3182)
 AUTHORS Rydkina,E., Roux,V., Raoult,D. and Tarasevich,I.
 TITLE Direct Submission
 JOURNAL Submitted (14-JAN-1999) Unite Des Rickettsies, Faculte De Medecine,

```
FEATURES
  source
    27 Boulevard Jean Moulin, Marseille 13385, France
    Location/Qualifiers
      1. .3182
        /organism="Rickettsia sp. Dns28"
        /strain="Dns28"
        /db_xref="taxon:114649"
        /note="isolated in the former USSR"
      <1. .>3182
        /gene="ompA"
        /gene="ompA"
        /note="outer membrane protein"
      <1. .>3182
        /codon_start=3
        /transl_table=11
        /product="OmpA"
        /protein_id="AAF34347.1"
        /db_xref="GI:6979180"
        /translation="ITLQAGGSLAANNIDFGVSGTLEFNGPLDGGNTIPYFKGAIA
        NGNALNVNTKSLTAYHSTIGTVAEINIGAGNLPADASAGDVTILNAQDINFGAPD
        STLALNLTGVGVKNILLAADLVAPGANAGDVVFDGGVNGNLNIGSVAGTARNIGDGG
        GNKFTLLIYNAVITDDVNLEGIQNVLINNADFTSSAFNAGAIQINDATYITIDAN
        NGNLVPAGNIQFAHADAKLILQNSGNDRTITLGANIDPDNDYEGIVTLNVSATGKK
        LTIAGKTLGGAHKLQAIIVFKAGDVSAGTTFNTNVVLDITGQLELGATFANVLL
        NDVOLTGTGNGIGFLDNKAGTFTLNNNVAGAVONTGTTNSGTLVLGASNLNS
        VNGIAMLKVGAGNVTIAKGDVYKIGEIQGTGNTLILPANFLGINSIKTSGOALKLN
        FTNGSVGVGTANSVGDITAGTATSFASVNAKGTATLGTTTFADTFTNTGAVT
        LAKGSITFNKNTATSVANSATINFGNSLAFNSNITGSGTTLTLAGANQVKTGTS
        FTDMLTNTFDGAKSGNIIKSGNTLDSGSTLALVATFADTATDIDDFEFGG
        VISAETAGLPTTEENVKTIINNDREFVDFTDASTLTLFAEDATDIDDFEFGG
        PLANTPNAANTKKSLEMDAPNGSDAQAFNNFGLMTPLQPADATTHLMQDVVKPSD
        TIAAVNNOVASNITSSNITFALNARDKVOAANKGPVSSGDEMDAKFGWISPFVNGA
        TQKMCNNTSGYKSDTGTGTGFDGLVSDLLGLAYTRADDTIKLKNKTGDKNVES
        NIYSLGYNVPEYENLFAEIAFYSDNKRISRRVIATALETVGTGANGKYSSEY
        TQLMAGTYMMPENILPLAGLRYSTIKQYKGTGTYQNLTVKGNKNTFQGLL
        GAKVSNINVEIPLTPELYAMVDYAFKNKPAIDARLQGMATPLPTNSFKQSKTSFD
        VGVGTAKHKMMEYGINVDNIGSKYFAQQ"
      BASE COUNT      1064 a      535 c      670 g      913 t
      ORIGIN

    Query Match      1.4%; Score 20; DB 2; Length 3182;
    Best Local Similarity 100.0%; Pred. No. 6.6;
    Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

  QY 1373 aggcattgccgaataaataat 1392
      |||||||||||||||
  Db 1618 AGGCTATTGCCGAATAAT 1599

RESULT 10
AF120020S1/c
LOCUS      AF120020S1      3182 bp      DNA      BCT      16-FEB-2000
DEFINITION Rickettsia sp. Dns14 OmpA (ompA) gene, partial cds.
ACCESSION  AF120020
VERSION     AF120020.1 GI:6979182
KEYWORDS
SEGMENT
SOURCE
  ORGANISM
    Rickettsia sp. Dns14.
    Rickettsia sp. Dns14
    Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
    Rickettsiaceae; Rickettsieae; Rickettsia; spotted fever group.
  1 (bases 1 to 3182)
  Rydkina,E., Roux,V., Fetisova,N., Rudakov,N., Gafarova,M.,
  Tarasevich,I. and Raoult,D.
  Detection of New Spotted Fever Group Rickettsiae in Ticks Collected
  in the Former USSR
  JOURNAL
  REFERENCE
  2 (bases 1 to 3182)
  Rydkina,E., Roux,V., Raoult,D. and Tarasevich,I.
  Direct Submission
  TITLE
  JOURNAL
  SUBMITTED (14-JAN-1999) Unite des Rickettsies, Faculte de Medecine,
  27 Boulevard Jean Moulin, Marseille 13385, France
  FEATURES
    Location/Qualifiers
      1. .3182
        /organism="Rickettsia sp. Dns14"
        /strain="Dns14"
        /db_xref="taxon:114648"
        /note="isolated in the former USSR"
      <1. .>3182
        /gene="ompA"
        /gene="ompA"
        /note="outer membrane protein"
      <1. .>3182
        /codon_start=3
        /transl_table=11
        /product="OmpA"
        /protein_id="AAF34348.1"
        /db_xref="GI:6979184"
        /translation="ITLQAGGSLAANNIDFGVSGTLEFNGPLDGGNTIPYFKGAIA
        NGNALNVNTKSLTAYHSTIGTVAEINIGAGNLPADASAGDVTILNAQDINFGAPD
        STLALNLTGVGVKNILLAADLVAPGANAGDVVFDGGVNGNLNIGSVAGTARNIGDGG
        GNKFTLLIYNAVITDDVNLEGIQNVLINNADFTSSAFNAGAIQINDATYITIDAN
        NGNLVPAGNIQFAHADAKLILQNSGNDRTITLGANIDPDNDYEGIVTLNVSATGKK
        LTIAGKTLGGAHKLQAIIVFKAGDVSAGTTFNTNVVLDITGQLELGATFANVLL
        NDVOLTGTGNGIGFLDNKAGTFTLNNNVAGAVONTGTTNSGTLVLGASNLNS
        VNGIAMLKVGAGNVTIAKGDVYKIGEIQGTGNTLILPANFLGINSIKTSGOALKLN
        FTNGSVGVGTANSVGDITAGTATSFASVNAKGTATLGTTTFADTFTNTGAVT
        LAKGSITFNKNTATSVANSATINFGNSLAFNSNITGSGTTLTLAGANQVKTGTS
        FTDMLTNTFDGAKSGNIIKSGNTLDSGSTLALVATFADTATDIDDFEFGG
        VISAETAGLPTTEENVKTIINNDREFVDFTDASTLTLFAEDATDIDDFEFGG
        PLANTPNAANTKKSLEMDAPNGSDAQAFNNFGLMTPLQPADATTHLMQDVVKPSD
        TIAAVNNOVASNITSSNITFALNARDKVOAANKGPVSSGDEMDAKFGWISPFVNGA
        TQKMCNNTSGYKSDTGTGTGFDGLVSDLLGLAYTRADDTIKLKNKTGDKNVES
        NIYSLGYNVPEYENLFAEIAFYSDNKRISRRVIATALETVGTGANGKYSSEY
        TQLMAGTYMMPENILPLAGLRYSTIKQYKGTGTYQNLTVKGNKNTFQGLL
        GAKVSNINVEIPLTPELYAMVDYAFKNKPAIDARLQGMATPLPTNSFKQSKTSFD
        VGVGTAKHKMMEYGINVDNIGSKYFAQQ"
      BASE COUNT      1063 a      533 c      673 g      913 t
      ORIGIN

    Query Match      1.4%; Score 20; DB 2; Length 3182;
    Best Local Similarity 100.0%; Pred. No. 6.6;
    Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

  QY 1373 aggcattgccgaataaataat 1392
      |||||||||||||||
  Db 1618 AGGCTATTGCCGAATAAT 1599

RESULT 11
AF120022S2/c
LOCUS      AF120022S2      3182 bp      DNA      BCT      16-FEB-2000
DEFINITION Rickettsia sp. RpA4 OmpA (ompA) gene, partial cds.
ACCESSION  AF120023
VERSION     AF120023.1 GI:6979188
KEYWORDS
SEGMENT
SOURCE
  ORGANISM
    Rickettsia sp. RpA4.
    Rickettsia sp. RpA4
    Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
    Rickettsiaceae; Rickettsieae; Rickettsia; spotted fever group.
  1 (bases 1 to 3182)
  Rydkina,E., Roux,V., Fetisova,N., Rudakov,N., Gafarova,M.,
  Tarasevich,I. and Raoult,D.
  Detection of New Spotted Fever Group Rickettsiae in Ticks Collected
  in the Former USSR
  JOURNAL
  REFERENCE
  2 (bases 1 to 3182)
  Rydkina,E., Roux,V., Raoult,D. and Tarasevich,I.
  Direct Submission
  TITLE
  JOURNAL
  SUBMITTED (14-JAN-1999) Unite des Rickettsies, Faculte de Medecine,
  27 boulevard Jean Moulin, Marseille 13385, France
  FEATURES
    Location/Qualifiers
      1. .3182
        /organism="Rickettsia sp. RpA4"
        /strain="Dns14"
        /db_xref="taxon:114648"
        /note="isolated in the former USSR"
      <1. .>3182
        /gene="ompA"
        /gene="ompA"
        /note="outer membrane protein"
      <1. .>3182
        /codon_start=3
        /transl_table=11
        /product="OmpA"
        /protein_id="AAF34348.1"
        /db_xref="GI:6979184"
        /translation="ITLQAGGSLAANNIDFGVSGTLEFNGPLDGGNTIPYFKGAIA
        NGNALNVNTKSLTAYHSTIGTVAEINIGAGNLPADASAGDVTILNAQDINFGAPD
        STLALNLTGVGVKNILLAADLVAPGANAGDVVFDGGVNGNLNIGSVAGTARNIGDGG
        GNKFTLLIYNAVITDDVNLEGIQNVLINNADFTSSAFNAGAIQINDATYITIDAN
        NGNLVPAGNIQFAHADAKLILQNSGNDRTITLGANIDPDNDYEGIVTLNVSATGKK
        LTIAGKTLGGAHKLQAIIVFKAGDVSAGTTFNTNVVLDITGQLELGATFANVLL
        NDVOLTGTGNGIGFLDNKAGTFTLNNNVAGAVONTGTTNSGTLVLGASNLNS
        VNGIAMLKVGAGNVTIAKGDVYKIGEIQGTGNTLILPANFLGINSIKTSGOALKLN
        FTNGSVGVGTANSVGDITAGTATSFASVNAKGTATLGTTTFADTFTNTGAVT
        LAKGSITFNKNTATSVANSATINFGNSLAFNSNITGSGTTLTLAGANQVKTGTS
        FTDMLTNTFDGAKSGNIIKSGNTLDSGSTLALVATFADTATDIDDFEFGG
        VISAETAGLPTTEENVKTIINNDREFVDFTDASTLTLFAEDATDIDDFEFGG
        PLANTPNAANTKKSLEMDAPNGSDAQAFNNFGLMTPLQPADATTHLMQDVVKPSD
        TIAAVNNOVASNITSSNITFALNARDKVOAANKGPVSSGDEMDAKFGWISPFVNGA
        TQKMCNNTSGYKSDTGTGTGFDGLVSDLLGLAYTRADDTIKLKNKTGDKNVES
        NIYSLGYNVPEYENLFAEIAFYSDNKRISRRVIATALETVGTGANGKYSSEY
        TQLMAGTYMMPENILPLAGLRYSTIKQYKGTGTYQNLTVKGNKNTFQGLL
        GAKVSNINVEIPLTPELYAMVDYAFKNKPAIDARLQGMATPLPTNSFKQSKTSFD
        VGVGTAKHKMMEYGINVDNIGSKYFAQQ"
      BASE COUNT      1063 a      533 c      673 g      913 t
      ORIGIN
```

[illegible]

```
/product="OmpA"
/db_xref="AAC35176.1"
/translation="IILQAGGSLDANNIDFGARSTLEFNGPLDGGGNAIPYVFKGAIA
NGNAILNVNTKLLTAYHLTIGTVAEINIGAGNLFDAIDASAGDVTILNAQDIHFRALD
SALVLSNLTCGVNNILLAAIDLVPQVDEGTVPVFDGGVNGNLGNSVAGAAARNIGDVG
GNKNTLLIYNAVYITDDVNLEGIQNVLLNNADFTSTAFNAGTIQINDATYTTIDAN
KDAVOLTOGTNGIGFLDFNAKNGVTILNNNVNAGTVKNTGTTNGTTLVLGASNLNR
LTIAGKTFGGAAHKLQAIIVFKGEGDFGTAGTTFNTNIVLDTIGOLELGATTANVVL
KDAVOLTOGTNGIGFLDFNAKNGVTILNNNVNAGTVKNTGTTNGTTLVLGASNLNR
VINGIAMLKVGAGNVTIAGKNGVIGIQTGNTTLTLPFAHFKLTGTSINTGGOALKLN
FMNGSVGVGTAAVSDITTAGTSFASVNAKGTATLGTTSTFAHFTNTGAVT
LAKGSIITFAKNTATSVANSATINFGNSLAFNSITGTTLTGANGVYTTGTGS
FTDTLTLNTFDGAAGSGNLIKSGSTLDLSGVSNLAVTATNFMNNSPDTKYT
VISATAGLTKPTKENVKITINNDRFVDTFADSTLTLFAEDIAAGVIDEDFAPGG
PLANTPNAANIKKSLLEMDAPGSDAQAFNFGMLTFLQEDATTHLMQDVYFVPSD
TIAAVNNQVVASNISNITARNMDKVOAGNKGVPSSGDEMDAKFGAWISFPVGN
TOKMNSISGVKSDTGTGTFDGVSDLVGLAYTRADTDIKLNNKTGDKNKVES
NIYSLYGLIYSPYENLVEAIASDNKIRSKRVYATTLTETVGTQFANGYKSES
TGOLMAGTYIMSENINLTPLAGLRYSTIKDKSYKETGTYQNTLVTKGNYNTEFGLL
GAKVSSNNVNEIIVLPELYAMVDYAFKNKVSADARLQGMATPLPTNSFKQKTSFD
VGCVGTAKHKMMEYGINYDNTNIGSKYFAQQ"
BASE COUNT      1065 a      510 c      673 g      934 t
ORIGIN

Query Match      1.4%; Score 20; DB 2; Length 3182;
Best Local Similarity 100.0%; Pred. No. 6.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1373 aggtattgcgaataat 1392
|||||
Db 1618 AGGCTATTGCCGAATAAT 1599

RESULT 14
RCU83443/c
LOCUS      Rickettsia conorii romPA (ompA) gene, partial cds.
DEFINITION
ACCESSION  U83443
VERSION     U83443.1 GI:1778876
KEYWORDS   Rickettsia conorii.
SOURCE     Rickettsia conorii
ORGANISM   Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
            Rickettsiaceae; Rickettsiae; Rickettsia.
REFERENCE  Fournier,P.E., Roux,V. and Raoult,D.
AUTHORS    Phylogenetic analysis of spotted fever group rickettsiae by study
            of the outer surface protein romPA
TITLE      Int. J. Syst. Bacteriol. 48 Pt 3, 839-849 (1998)
JOURNAL    98404549
MEDLINE    2 (bases 1 to 3182)
REFERENCE  Raoult,D., Fournier,P.-E. and Roux,V.
AUTHORS    Direct Submission
TITLE      Submitted (31-DEC-1996) Faculte de Medecine, Unite des Rickettsies,
            CNRS EP J0054, 27 bd Jean Moulin, Cedex 5, Marseille, France
JOURNAL    CNRS EP J0054, 27 bd Jean Moulin, Cedex 5, Marseille, France
FEATURES    Location/Qualifiers
            source
            1..3182
            /organism="Rickettsia conorii"
            /strain="M1"
            /db_xref="taxon:781"
            1..3182
            /gene="ompA"
            <1..>3182
            /gene="ompA"
            /codon_start=3
            /transl_table=11
            /product="OmpA"
            /protein_id="AAC35179.1"
            /db_xref="GI:1778877"
            /translation="VTIQAGGSLDANNIDFGARSTLEFNGPLDGGGNAIPYVFKGAIA
            NGNAILNVNTKLLTAYHLTIGTVAEINIGAGNLFDAIDASAGDVTILNAQDIHFRALD
            SALVLSNLTCGVNNILLAAIDLVPQVDEGTVPVFDGGVNGNLGNSVAGAAARNIGDVG
            GNKNTLLIYNAVYITDDVNLEGIQNVLLNNADFTSTAFNAGTIQINDATYTTIDAN
            KDAVOLTOGTNGIGFLDFNAKNGVTILNNNVNAGTVKNTGTTNGTTLVLGASNLNR
            LTIAGKTFGGAAHKLQAIIVFKGEGDFGTAGTTFNTNIVLDTIGOLELGATTANVVL
            KDAVOLTOGTNGIGFLDFNAKNGVTILNNNVNAGTVKNTGTTNGTTLVLGASNLNR
            VINGIAMLKVGAGNVTIAGKNGVIGIQTGNTTLTLPFAHFKLTGTSINTGGOALKLN
            FMNGSVGVGTAAVSDITTAGTSFASVNAKGTATLGTTSTFAHFTNTGAVT
            LAKGSIITFAKNTATSVANSATINFGNSLAFNSITGTTLTGANGVYTTGTGS
            FTDTLTLNTFDGAAGSGNLIKSGSTLDLSGVSNLAVTATNFMNNSPDTKYT
            VISATAGLTKPTKENVKITINNDRFVDTFADSTLTLFAEDIAAGVIDEDFAPGG
            PLANTPNAANIKKSLLEMDAPGSDAQAFNFGMLTFLQEDATTHLMQDVYFVPSD
            TIAAVNNQVVASNISNITARNMDKVOAGNKGVPSSGDEMDAKFGAWISFPVGN
            TOKMNSISGVKSDTGTGTFDGVSDLVGLAYTRADTDIKLNNKTGDKNKVES
            NIYSLYGLIYSPYENLVEAIASDNKIRSKRVYATTLTETVGTQFANGYKSES
            TGOLMAGTYIMSENINLTPLAGLRYSTIKDKSYKETGTYQNTLVTKGNYNTEFGLL
            GAKVSSNNVNEIIVLPELYAMVDYAFKNKVSADARLQGMATPLPTNSFKQKTSFD
            VGCVGTAKHKMMEYGINYDNTNIGSKYFAQQ"
            510 c      673 g      934 t
```

```
SALVLSNLTCGVNNILLAAIDLVPQVDEGTVPVFDGGVNGNLGNSVAGAAARNIGDVG
GNKNTLLIYNAVYITDDVNLEGIQNVLLNNADFTSTAFNAGTIQINDATYTTIDAN
NGNAILPAGNIKFAHADQAIIVFKGEGDFGTAGTTFNTNIVLDTIGOLELGATTANVVL
LTIAGKTFGGAAHKLQAIIVFKGEGDFGTAGTTFNTNIVLDTIGOLELGATTANVVL
KDAVOLTOGTNGIGFLDFNAKNGVTILNNNVNAGTVKNTGTTNGTTLVLGASNLNR
VINGIAMLKVGAGNVTIAGKNGVIGIQTGNTTLTLPFAHFKLTGTSINTGGOALKLN
FMNGSVGVGTAAVSDITTAGTSFASVNAKGTATLGTTSTFAHFTNTGAVT
LAKGSIITFAKNTATSVANSATINFGNSLAFNSITGTTLTGANGVYTTGTGS
FTDTLTLNTFDGAAGSGNLIKSGSTLDLSGVSNLAVTATNFMNNSPDTKYT
VISATAGLTKPTKENVKITINNDRFVDTFADSTLTLFAEDIAAGVIDEDFAPGG
PLANTPNAANIKKSLLEMDAPGSDAQAFNFGMLTFLQEDATTHLMQDVYFVPSD
TIAAVNNQVVASNISNITARNMDKVOAGNKGVPSSGDEMDAKFGAWISFPVGN
TOKMNSISGVKSDTGTGTFDGVSDLVGLAYTRADTDIKLNNKTGDKNKVES
NIYSLYGLIYSPYENLVEAIASDNKIRSKRVYATTLTETVGTQFANGYKSES
TGOLMAGTYIMSENINLTPLAGLRYSTIKDKSYKETGTYQNTLVTKGNYNTEFGLL
GAKVSSNNVNEIIVLPELYAMVDYAFKNKVSADARLQGMATPLPTNSFKQKTSFD
VGCVGTAKHKMMEYGINYDNTNIGSKYFAQQ"
BASE COUNT      1064 a      511 c      674 g      933 t
ORIGIN

Query Match      1.4%; Score 20; DB 2; Length 3182;
Best Local Similarity 100.0%; Pred. No. 6.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1373 aggtattgcgaataat 1392
|||||
Db 1618 AGGCTATTGCCGAATAAT 1599

RESULT 15
RCU83448/c
LOCUS      Rickettsia conorii romPA (ompA) gene, partial cds.
DEFINITION
ACCESSION  U83448
VERSION     U83448.1 GI:1778878
KEYWORDS   Rickettsia conorii.
SOURCE     Rickettsia conorii
ORGANISM   Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
            Rickettsiaceae; Rickettsiae; Rickettsia.
REFERENCE  Fournier,P.E., Roux,V. and Raoult,D.
AUTHORS    Phylogenetic analysis of spotted fever group rickettsiae by study
            of the outer surface protein romPA
TITLE      Int. J. Syst. Bacteriol. 48 Pt 3, 839-849 (1998)
JOURNAL    98404549
MEDLINE    2 (bases 1 to 3182)
REFERENCE  Raoult,D., Fournier,P.-E. and Roux,V.
AUTHORS    Direct Submission
TITLE      Submitted (31-DEC-1996) Faculte de Medecine, Unite des Rickettsies,
            CNRS EP J0054, 27 bd Jean Moulin, Cedex 5, Marseille, France
JOURNAL    CNRS EP J0054, 27 bd Jean Moulin, Cedex 5, Marseille, France
FEATURES    Location/Qualifiers
            source
            1..3182
            /organism="Rickettsia conorii"
            /strain="Moroccan"
            /db_xref="taxon:781"
            1..3182
            /gene="ompA"
            <1..>3182
            /gene="ompA"
            /codon_start=3
            /transl_table=11
            /product="romPA"
            /protein_id="AAC35184.1"
            /db_xref="GI:1778879"
            /translation="VTIQAGGSLDANNIDFGARSTLEFNGPLDGGGNAIPYVFKGAIA
            NGNAILNVNTKLLTAYHLTIGTVAEINIGAGNLFDAIDASAGDVTILNAQDIHFRALD
            SALVLSNLTCGVNNILLAAIDLVPQVDEGTVPVFDGGVNGNLGNSVAGAAARNIGDVG
            GNKNTLLIYNAVYITDDVNLEGIQNVLLNNADFTSTAFNAGTIQINDATYTTIDAN
            KDAVOLTOGTNGIGFLDFNAKNGVTILNNNVNAGTVKNTGTTNGTTLVLGASNLNR
            LTIAGKTFGGAAHKLQAIIVFKGEGDFGTAGTTFNTNIVLDTIGOLELGATTANVVL
            KDAVOLTOGTNGIGFLDFNAKNGVTILNNNVNAGTVKNTGTTNGTTLVLGASNLNR
```

VNCIAMLKVGACNVITLAKGNVKGTEQGTGNTNLTLPHPFKLTSGINTKGQALKL
FMNGSVSGVGTAAANSVDITTAGATSVFSSVNAKGTATLGTTTSAFTNTTNGAT
LARGSTISPAKNTTATSVFANSTIFENGSLANPSNTGTTTLTCAQVNTTNGTGS
FYDTLPLNTFDCAAKSNGKILKSGSTDLGSGVNLAVVTATNPMNNISPDRTKY
VISAETPAGGTLKPKENVKLITNNDNRPVDTFADASTLTFAEDIAAGYIDEDPAGT
PLANTANAKINIKKLELMEDNAPSADQAFNFMGLMTPLQEDATLHLMDDVXPSS
TIAAVNVNVASINSNTALNARDKVQAGKSGPSGSDMDKAGWISPFVNGA
TORMNSISGYKSDTGTGIDGDFVSDLVGLIYVRAOTDLTKLNNKTGOKNKVES
NIYSLGLYSPVENLFEVATASDNKISRSRVLATTLTGVYGTANGKVKSESY
TGQLMAGTYTVMSENI NLTPLAGLRYSTIKDKSYCTGTQYNLPVGNKNYNTPDGLL
KAGVSSINI NVSEIPLTAVWDYATKFNYSADARLQQTQNTPLTNGKNSQKSTGFL
VGVGYTAKHMMEYGINDTNIGSKYTAQ*

512 c 675 a 933 T

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 9, 2000, 21:26:38 ; Search time 66.13 Seconds
(without alignments)
5455.572 Million cell updates/sec

Title: US-09-214-679-1

Perfect score: 1442

Sequence: 1 ccgggaactccatgtgcc.....aatgcaattcatttgatcc 1442

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 311585 seqs, 125096042 residues

Word size : 0

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	DB	ID	Description
1	1442	100.0	1442	1	V10449	K. oxytoca R-speci
2	18	1.2	1263	1	N82205	Nad B gene encodin
3	18	1.2	1442	1	V10449	K. oxytoca R-speci
4	18	1.2	1724	1	Q03361	The nadA nucleotid
5	18	1.2	1724	1	Q03363	The nadA nucleotid
6	17	1.2	26	1	Q49599	lacS gene C-termin
7	17	1.2	26	1	Q49600	lacS gene C-termin
8	17	1.2	26	1	Q51468	Primer 2. Stable i
9	17	1.2	26	1	Q51469	Primer 3. Stable i
10	17	1.2	1196	1	V20812	Human haematopoiet
11	17	1.2	1501	1	V00522	HIV-1 breakthrough
12	17	1.2	1512	1	V00521	HIV-1 breakthrough
13	17	1.2	1883	1	X20225	Enterococcus faeca
14	17	1.2	2019	1	X20306	Borrelia burgdorfe
15	17	1.2	3800	1	X28141	CCOAMF promoter,
16	17	1.2	3963	1	X20224	Enterococcus faeca
17	17	1.2	3963	1	X20226	Enterococcus faeca
18	17	1.2	4305	1	X20275	Borrelia burgdorfe
19	17	1.2	4312	1	X20274	Borrelia burgdorfe
20	17	1.2	5558	1	X12948	Enterococcus faeca
21	17	1.2	6444	1	V74544	Staphylococcus aur
22	17	1.2	7227	1	N60194	Sequence encoding
23	17	1.2	9813	1	X13059	Enterococcus faeca
24	16	1.1	99	1	Q21267	PND EE322-1. New a
25	16	1.1	102	1	Q21239	PND EE244-2. New a
26	16	1.1	102	1	Q21236	PND EE229-2. New a
27	16	1.1	102	1	Q21235	PND EE229-1. New a
28	16	1.1	102	1	Q21237	PND EE229-3. New a
29	16	1.1	102	1	Q21238	PND EE244-1. New a
30	16	1.1	102	1	Q21240	PND EE244-3. New a
31	16	1.1	102	1	Q21371	PND EE1476-1. New a
32	16	1.1	105	1	Q21309	PND EE505-2. New a
33	16	1.1	105	1	Q21310	PND EE505-3. New a

34	16	1.1	105	1	Q21311	PND EE507-1. New a
35	16	1.1	105	1	Q21221	PND EE179-3. New a
36	16	1.1	105	1	Q21258	PND EE310-2. New a
37	16	1.1	105	1	Q21326	PND EE520-1. New a
38	16	1.1	105	1	Q21327	PND EE520-2. New a
39	16	1.1	105	1	Q21328	PND EE520-3. New a
40	16	1.1	105	1	Q21372	PND EE5032-1. New a
41	16	1.1	105	1	Q21058	PND EE37-2. New a
42	16	1.1	105	1	Q21329	PND EE28-1. New a
43	16	1.1	105	1	Q21061	PND EE69-1. New a
44	16	1.1	105	1	Q21330	PND EE528-2. New a
45	16	1.1	105	1	Q21194	PND EE69-2. New a

ALIGNMENTS

RESULT 1

```
V10449
ID V10449 standard; DNA; 1442 BP.
AC V10449;
DT 16-JUN-1998 (first entry)
DE K. oxytoca R-specific amidohydrolase genomic DNA.
KW R-specific amidohydrolase; hydrolysis; racemate; nitrogen source;
   3,3,3-trifluoro-2-hydroxy-2-methyl propionamide; isoform; ds.
OS Klebsiella oxytoca.
FH Key Location/Qualifiers
   CDS 197..1183
      /tag=a
      /product=
      /note="R-specific amidohydrolase"
FT
FT
FT
PN WO9801568-A2.
PD 15-JAN-1998.
PF 10-JUL-1997; E03670.
PR 03-MAR-1997; CH-000500.
PR 10-JUL-1996; CH-001723.
PA (LONZ ) LONZA AG.
PI Brieden W, Naughton A, Robins K, Shaw N, Tinschert A,
   Zimmermann T;
DR WPI: 98-101063/09.
DR P-PSDB: W40263.
PT Tri:fluoro-2-hydroxy-2-methyl-propionic acid enantiomer preparation
   - by stereoselective hydrolysis of corresponding racemic amide using
   microorganism or derived enzyme, used as drug intermediate
PS Claim 7a; Page 32-34; 68pp; German.
CC This DNA sequence encodes a R-specific amidohydrolase isolated from
   Klebsiella oxytoca strain P81 which allows the microorganism to
   utilize 3,3,3-trifluoro-2-hydroxy-2-methyl propionamide as its sole
   nitrogen source. This amidohydrolase is used in a process for preparing
   (R)-isoforms of 3,3,3-trifluoro-2-hydroxy-2-methyl propionamide which
   is cheaper than prior art optical resolution of the racemate using
   dimethoxy strychnine or (S)-(-)- alpha -methylbenzylamine.
SQ Sequence 1442 BP; 385 A; 350 C; 370 G; 337 T;
```

Query Match 100.0%; Score 1442; DB 1; Length 1442;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1442; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	ccgggaactccatgtgccgtgattgctgcagaggattggtgatccagcggg	60
DB	1	CCCGGGAATCCATGTGGCGGATCTGTCGAGCAGGATATGGCGATGTCAGCGG	60
QY	61	ccgcacagcgtgtgcggttaataagcctgtgttagaaagcgtacccaacaac	120
DB	61	CCGCACAGCGCTGTCGGGTAATGATTAAGCGCTGTTGTAGAACCGCTACCCACAA	120
QY	121	agctctctgatgatcttttaataagcgtcgtcatctggtctgttaactaaacgtataaatt	180
DB	121	AGCTCTCTGATGATCTTTTAATGCGTCGTCATCTGGCTCTGTAACGCTATAAATT	180
QY	181	acgtgagataacacatgaaatggttgggaagaatccattatggccaacgcggtgttg	240

PA (LONZ) LONZA AG.
PI Brieden W, Naughton A, Robins K, Shaw N, Tinschert A,
PI Zimmermann T;
DR WPI; 98-101063/09.
DR P-PSDB; W40263.
PT Tri:fluoro-2-hydroxy-2-methyl-propionic acid enantiomer preparation
PT - by stereoselective hydrolysis of corresponding racemic amide using
PT microorganism or derived enzyme, used as drug intermediate
PS Claim 7a; Page 32-34; 68pp; German.
CC This DNA sequence encodes a R-specific amidohydrolase isolated from
CC Klebsiella oxytoca strain PRS1 which allows the microorganism to
CC utilise 3,3,3-trifluoro-2-hydroxy-2-methyl propionamide as its sole
CC nitrogen source. This amidohydrolase is used in a process for preparing
CC (R)-isomers of 3,3,3-trifluoro-2-hydroxy-2-methyl propionamide which
CC is cheaper than prior art optical resolution of the racemate using
CC dimethoxy strychnine or (S)-(-)- alpha -methylbenzylamine.
SQ Sequence 1442 BP; 385 A; 350 C; 370 G; 337 T;

Query Match 1.2%; Score 18; DB 1; Length 1442;
Best Local Similarity 100.0%; Pred. No. 7.4;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 61 cgcacagcgtgtgcgg 78
DB 78 CGCACAGCGCTGTCGG 61

RESULT 4
Q03361/C
ID Q03361 standard; DNA; 1724 BP.
AC Q03361;
DE 24-JUL-1990 (first entry)
DE The nadA nucleotide sequence used in effecting prodn. of quinolinic acid
KW Quinolinic acid production; nadA; nadB; pharmaceutical intermediate;
KW Plant protection agent intermediate; nadA; nadB; pharmaceutical intermediate;
KW L-aspartate oxidase; Escherichia coli W4546; quinolinic acid synthase;
PN DE3826040-A.
PD 08-FEB-1990.
PF 30-JUL-1988; 826040.
PR Laufer A, Gassen HG, Flachmann R, Kunz N;
PA (RUTG) Rutgerswerke Ag.
PI Laufer A, Gassen HG, Flachmann R, Kunz N;
DR WPI; 90-045655/07.
PT Fermentative prodn. of quinolinic acid -
PT using genetically transformed microorganisms contg. nadA and
PT nadB genes
PS Disclosure; p; German.
CC It codes for quinolinic acid synthase. Also claimed are nadB which codes
CC for L-aspartate oxidase and genetically transformed microorganisms contg.
CC plasmids bearing nadA and nadB in a copy ratio of 200:50 or 200:8.
CC Transformation of E. coli W4546 successively with PQAB 520 and PQAA 168
CC produced a transformant with a 200:50 copy ratio, giving quinolinic acid
CC in yields which were ca. 7.5-9.7 times those obtained using W4546. The
CC microorganisms give high yields in a reproducible manner. Quinolinic
CC acid is an intermediate for pharmaceuticals and plant protection agents.
SQ Sequence 1724 BP; 408 A; 432 C; 470 G; 414 T;

Query Match 1.2%; Score 18; DB 1; Length 1724;
Best Local Similarity 100.0%; Pred. No. 7.5;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 459 tcattgtcaatgctgcgg 476
DB 976 TCATGTCATGCGCGCG 959

RESULT 5
Q03363/C
ID Q03363 standard; DNA; 1724 BP.
AC Q03363;
DE 24-JUL-1990 (first entry)

DE The nadA nucleotide sequence used in effecting prodn. of quinolinic acid
KW Quinolinic acid production; nadA; nadB; pharmaceutical intermediate;
KW Plant protection agent intermediate; nadA; nadB; pharmaceutical intermediate;
KW L-aspartate oxidase; Escherichia coli W4546.
PN DE3826040-A.
PD 08-FEB-1990.
PF 30-JUL-1988; 826040.
PR Laufer A, Gassen HG, Flachmann R, Kunz N;
PA (RUTG) Rutgerswerke Ag.
PI Laufer A, Gassen HG, Flachmann R, Kunz N, Seifert J;
DR WPI; 90-045656/07.
PT Fermentative prodn. of quinolinic acid -
PT using genetically transformed microorganisms contg. nadA and
PT nadB genes
PS Disclosure; p; German.
CC It codes for quinolinic acid synthase. Also claimed are nadB which codes
CC for L-aspartate oxidase and genetically transformed microorganisms contg.
CC plasmids bearing nadA and nadB in a copy ratio of 50:200, 8:50 or
CC 8:200. Isolated chromosomal DNA was digested with HindII to obtain 6 and
CC 8 kb fragments, which were inserted into pBR322 and cloned in E. coli
CC NK6042 to obtain plasmid PQAB510. This was digested with HindIII and NruI
CC to obtain a 3.2 kb fragment, which was inserted in pBR322 to produce PQAB
CC 520. This was digested with SspI and NruI to obtain a 2.5 kb fragment,
CC which was inserted in pUC18 to obtain PQAB203. This was digested with
CC PstI and BamHI to give a 2.5 kb fragment, which was inserted into pBR234
CC to give the nadB-contg. plasmid PQAB104. The same chromosomal DNA was
CC digested with BamHI, inserted into pLG339 and cloned in E. coli RFI to
CC obtain PQAA12. This was digested with AluI, ligated with HincII-cleaved
CC pUC18 and cloned in E. coli 431 to give PQAA166. This was digested with
CC BamII and PstI to isolate a 1.4 kb insert, which was inserted into PstI/
CC EcoRI-digested pBR322 to obtain the nadA-contg. plasmid PQAA169.
CC Transformation of E. coli W4546 successively with PQAB 104 and PQAA 169
CC produced a transformant with a 50:200 copy ratio, giving quinolinic acid
CC in yields which were ca. 2.6-2.8 times those obtained using W4546. The
CC microorganisms give high yields in a reproducible manner. Quinolinic
CC acid is an intermediate for pharmaceuticals and plant protection agents.
SQ Sequence 1724 BP; 408 A; 432 C; 470 G; 414 T;

Query Match 1.2%; Score 18; DB 1; Length 1724;
Best Local Similarity 100.0%; Pred. No. 7.5;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 459 tcattgtcaatgctgcgg 476
DB 976 TCATGTCATGCGCGCG 959

RESULT 6
Q49599/C
ID Q49599 standard; DNA; 26 BP.
AC Q49599;
DE 27-APR-1994 (first entry)
DE lac gene C-terminal fragment PCR primer 2.
KW Polymerase chain reaction; amplification; donor plasmid;
KW construction; ss.
OS Synthetic.
PN EP-564965-A.
PD 13-OCT-1993.
PF 31-MAR-1993; 105303.
PR 07-APR-1992; EP-105973.
PA (NEST) SOC PROD NESTLE SA.
PI Knol J, Marciset O, Mollet B;
DR WPI; 93-322227/41.
PT Integral gene expression in food-grade microorganisms - by
PT insertion of a promoter-less gene into an operon in front of an
PT essential cistron
PS Example 1; Page 6; 24pp; English.
CC The sequence is that of a PCR primer which was used in the
CC amplification of a ca. 900 bp long fragment, pBM37, contg. the C-terminal
CC end of lacS from E. coli W4546.

CC NdeI site. The PCR amplification was performed as part of the
 CC construction of donor plasmids for the integration of the cat
 CC gene into the lac operon.
 SQ Sequence 26 BP; 4 A; 5 C; 3 G; 14 T;

Query Match 1.2%; Score 17; DB 1; Length 26;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 185 ggagaataacatatgaa 201
 |||||
 Db 23 GGAGATAACATATGAA 7

RESULT 7
 Q49600
 ID Q49600 standard; DNA; 26 BP.
 AC Q49600;
 DT 27-APR-1994 (first entry)
 DE lacS gene C-terminal fragment PCR primer 3.
 KW Polymerase chain reaction; amplification; donor plasmid;
 KW construction; ss.
 OS Synthetic.
 PN EP-564965-A.
 PD 13-OCT-1993.
 PF 31-MAR-1993; 105303.
 PR 07-APR-1992; EP-105973.
 PA (NEST) SOC PROD NESTLE SA.
 PI Knol J, Marciset O, Mollet B;
 DR WPI; 93-32227/41.
 PT Integral gene expression in food-grade microorganisms - by
 PT insertion of a promoter-less gene into an operon in front of an
 PT essential cistron
 PS Example 1; Page 6; 24pp; English.
 CC The sequence is that of a PCR primer which was used in the
 CC amplification of a ca. 700 bp long fragment, pBM38, contg. the C-terminal
 CC end of lacS from FspI linearised pD228. The PCR amplification was
 CC performed as part of the construction of donor plasmids for the
 CC integration of the cat gene into the lac operon.
 SQ Sequence 26 BP; 14 A; 3 C; 5 G; 4 T;

Query Match 1.2%; Score 17; DB 1; Length 26;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 185 ggagaataacatatgaa 201
 |||||
 Db 4 GGAGATAACATATGAA 20

RESULT 8
 Q51468/c
 ID Q51468 standard; DNA; 26 BP.
 AC Q51468;
 DT 19-MAY-1994 (first entry)
 DE Primer 2.
 KW Polymerase chain reaction; PCR; primer; amplify; stable; integration;
 KW homogenic; heterogenic; Streptococcus thermophilus; restriction site;
 KW donor plasmid; artificial; milk product; yoghurt; ss.
 OS Synthetic.
 PN EP-569604-A.
 PD 18-NOV-1993.
 PF 07-APR-1992; 105973.
 PR 07-APR-1992; EP-105973.
 PA (NEST) SOC PROD NESTLE SA.
 PI Knol J, Marciset O, Mollet B;
 DR WPI; 93-360915/46.
 PT Stable integration of homo-, heterogenic or artificial genes into
 PT streptococcus thermophilus genome - to produce an acidified milk
 PT prod., pref. yoghurt
 PS Example 2; Page 5; 9pp; English.

CC The sequences given in Q51467-72 are primers which were used in the
 CC stable integration of homo- or heterogenic genes into the
 CC Streptococcus thermophilus genome. These plasmids were used to
 CC create restriction sites within donor plasmids which allowed
 CC insertion of the required genes into the S. thermophilus genome. S.
 CC thermophilus transformed in this way may be used in the production
 CC of an artificial milk product, especially yoghurt. Integration is
 CC stable and follows a strict copy number control of the host genome.
 CC The integrated gene may be expressed without the need for direct
 CC selection, observable phenotype or adaptation of the growth medium.
 SQ Sequence 26 BP; 4 A; 5 C; 3 G; 14 T;

Query Match 1.2%; Score 17; DB 1; Length 26;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 185 ggagaataacatatgaa 201
 |||||
 Db 23 GGAGATAACATATGAA 7

RESULT 9
 Q51469
 ID Q51469 standard; DNA; 26 BP.
 AC Q51469;
 DT 19-MAY-1994 (first entry)
 DE Primer 3.
 KW Polymerase chain reaction; PCR; primer; amplify; stable; integration;
 KW homogenic; heterogenic; Streptococcus thermophilus; restriction site;
 KW donor plasmid; artificial; milk product; yoghurt; ss.
 OS Synthetic.
 PN EP-569604-A.
 PD 18-NOV-1993.
 PF 07-APR-1992; 105973.
 PR 07-APR-1992; EP-105973.
 PA (NEST) SOC PROD NESTLE SA.
 PI Knol J, Marciset O, Mollet B;
 DR WPI; 93-360915/46.
 PT Stable integration of homo-, heterogenic or artificial genes into
 PT streptococcus thermophilus genome - to produce an acidified milk
 PT prod., pref. yoghurt
 PS Example 2; Page 5; 9pp; English.

CC The sequences given in Q51467-72 are primers which were used in the
 CC stable integration of homo- or heterogenic genes into the
 CC Streptococcus thermophilus genome. These plasmids were used to
 CC create restriction sites within donor plasmids which allowed
 CC insertion of the required genes into the S. thermophilus genome. S.
 CC thermophilus transformed in this way may be used in the production
 CC of an artificial milk product, especially yoghurt. Integration is
 CC stable and follows a strict copy number control of the host genome.
 CC The integrated gene may be expressed without the need for direct
 CC selection, observable phenotype or adaptation of the growth medium.
 SQ Sequence 26 BP; 14 A; 3 C; 5 G; 4 T;

Query Match 1.2%; Score 17; DB 1; Length 26;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 185 ggagaataacatatgaa 201
 |||||
 Db 4 GGAGATAACATATGAA 20

RESULT 10
 V20812
 ID V20812 standard; cDNA to mRNA; 1196 BP.
 AC V20812;
 DT 16-JUL-1998 (first entry)
 DE Human haematopoietic stem cell growth factor coding sequence.
 KW Haematopoietic stem cell growth factor; SCGF; burst-promoting activity;
 KW BPA; granulocyte macrophage colony stimulating activity; gene therapy;

KW GPA; haematopoietic cell disorder; bone marrow inhibition; human; ss.
 OS Homo sapiens.
 FH Key
 FT 140. .877
 CDS Location/Qualifiers
 FT /*tag= a
 FT 1097..1102
 FT /*tag= b
 PN WO980869-A1.
 PD 05-MAR-1998.
 PE 27-AUG-1997; J02985.
 PR 07-JUL-1997; WO-J02349.
 PR 27-AUG-1996; JP-262252.
 PR 24-MAR-1997; JP-087242.
 PA (KYOW) KYOWA HAKKO KOGYO KK.
 PI Hiraoka A, Mio H, Sugimura A;
 DR WPI; 98-179383/16.
 DR P-PSDB; W53245.
 PT Haematopoietic stem cell growth factor - useful for, e.g. treatment
 PT and diagnosis of haematopoietic cell abnormalities and bone marrow
 PT inhibition
 PS Claim 14; Page 42-44; 85pp; Japanese.
 CC This sequence encodes the human haematopoietic stem cell growth factor
 CC (SCGF) of the invention. The polypeptide of the invention is of mammalian
 CC origin and has haematopoietic stem cell growth factor SCGF activity,
 CC including burst-promoting activity (BPA) and granulocyte macrophage
 CC colony stimulating activity (GPA). The products can be used for
 CC treatment, diagnosis and analysis of haematopoietic cell disorders and
 CC bone marrow inhibition, e.g. by cytotoxic anticancer agents such as
 CC 5-fluorouracil. The products can also be used for amplification of
 CC haematopoietic cells in vitro, e.g. for use in marrow grafting and gene
 CC therapy by insertion of SCGF gene using a suitable therapeutic vector.
 SQ Sequence 1196 BP; 260 A; 378 C; 377 G; 181 T;

Query Match 1.2%; Score 17; DB 1; Length 1196;
 Best Local Similarity 100.0%; Pred. No. 25;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 946 ttctctggccacgaatgg 962
 DB 964 TTCTCTGGCCACGAATGG 980
 |||||

RESULT 11
 V00522 ID V00522 standard; DNA; 1501 BP.
 AC V00522;
 DT 20-JUL-1998 (first entry)
 DE HIV-1 breakthrough isolate clone C15.3 gp120 polypeptide DNA.
 KW HIV-1; envelope protein; gp120; MN-rgp120; vaccine; AIDS; ss.
 OS Human immunodeficiency virus type 1.
 FH Key Location/Qualifiers
 FT CDS 1..1491
 FT /*tag= a
 PN WO9801564-A1.
 PD 15-JAN-1998.
 PE 03-JUL-1997; U09690.
 PR 08-JUL-1996; US-676737.
 PA (GETH) GENENTECH INC.
 PI Berman PW;
 DR WPI; 98-101059/09.
 DR P-PSDB; W37059.
 PT Human immunodeficiency virus gp120 sequences from vaccine
 PT breakthrough strains - useful in providing added protection against
 PT HIV not provided by original vaccine
 PS Claim 6; Page 112-114; 193pp; English.
 CC Clone C15.3 encodes a gp120 envelope polypeptide (see W37059) of
 CC a breakthrough isolate of HIV-1 obtained from an individual
 CC infected with HIV-1 through high risk activity while participating
 CC in Phase I or Phase 2 trials of MN-rgp120, a candidate recombinant
 CC gp120 HIV-1 vaccine. Nucleotide sequences (see V00517-30) for
 CC each of 7 breakthrough isolates. The envelope glycoprotein genes
 CC were obtained from proviral DNA using PCR (see V00533-38). All 7
 CC envelope glycoproteins possess sequences typical of subtype (clade)
 CC B viruses. The overall homology with MN-rgp120 is 69-80%. Use of
 CC the gp120 polypeptides from one or more of the breakthrough
 CC isolates, usually together with MN-rgp120, can provide protection
 CC against HIV strains that are sufficiently different from the
 CC vaccine strain (e.g. MN-rgp120) that the vaccine does not confer
 CC protection against those strains.
 SQ Sequence 1512 BP; 580 A; 249 C; 320 G; 363 T;

CC were obtained from proviral DNA using PCR (see V00533-38). All 7
 CC envelope glycoproteins possess sequences typical of subtype (clade)
 CC B viruses. The overall homology with MN-rgp120 is 69-80%. Use of
 CC the gp120 polypeptides from one or more of the breakthrough
 CC isolates, usually together with MN-rgp120, can provide protection
 CC against HIV strains that are sufficiently different from the
 CC vaccine strain (e.g. MN-rgp120) that the vaccine does not confer
 CC protection against those strains.
 SQ Sequence 1501 BP; 575 A; 244 C; 320 G; 362 T;

Query Match 1.2%; Score 17; DB 1; Length 1501;
 Best Local Similarity 100.0%; Pred. No. 25;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 780 atataggaccaggaggt 796
 DB 821 ATATAGGACCAGGAGT 837
 |||||

RESULT 12
 V00521 ID V00521 standard; DNA; 1512 BP.
 AC V00521;
 DT 20-JUL-1998 (first entry)
 DE HIV-1 breakthrough isolate clone C15.2 gp120 polypeptide DNA.
 KW HIV-1; envelope protein; gp120; MN-rgp120; vaccine; AIDS; ss.
 OS Human immunodeficiency virus type 1.
 FH Key Location/Qualifiers
 FT CDS 1..1500
 FT /*tag= a
 PN WO9801564-A1.
 PD 15-JAN-1998.
 PE 03-JUL-1997; U09690.
 PR 08-JUL-1996; US-676737.
 PA (GETH) GENENTECH INC.
 PI Berman PW;
 DR WPI; 98-101059/09.
 DR P-PSDB; W37058.
 PT Human immunodeficiency virus gp120 sequences from vaccine
 PT breakthrough strains - useful in providing added protection against
 PT HIV not provided by original vaccine
 PS Claim 6; Page 108-110; 193pp; English.
 CC Clone C15.2 encodes a gp120 envelope polypeptide (see W37058) of
 CC a breakthrough isolate of HIV-1 obtained from an individual
 CC infected with HIV-1 through high risk activity while participating
 CC in Phase I or Phase 2 trials of MN-rgp120, a candidate recombinant
 CC gp120 HIV-1 vaccine. Nucleotide sequences (see V00517-30) for
 CC each of 7 breakthrough isolates. The envelope glycoprotein genes
 CC were obtained from proviral DNA using PCR (see V00533-38). All 7
 CC envelope glycoproteins possess sequences typical of subtype (clade)
 CC B viruses. The overall homology with MN-rgp120 is 69-80%. Use of
 CC the gp120 polypeptides from one or more of the breakthrough
 CC isolates, usually together with MN-rgp120, can provide protection
 CC against HIV strains that are sufficiently different from the
 CC vaccine strain (e.g. MN-rgp120) that the vaccine does not confer
 CC protection against those strains.
 SQ Sequence 1512 BP; 580 A; 249 C; 320 G; 363 T;

Query Match 1.2%; Score 17; DB 1; Length 1512;
 Best Local Similarity 100.0%; Pred. No. 25;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 780 atataggaccaggaggt 796
 DB 821 ATATAGGACCAGGAGT 837
 |||||

RESULT 13
 X20225 ID X20225 standard; DNA; 1883 BP.

AC X20225;
 DT 20-APR-1999 (first entry)
 DE Enterococcus faecalis EF121 gene fragment.
 KW Enterococcus faecalis; infection; vaccine; immune response; diagnosis;
 KW detection; attenuation; antigenic; ss.
 OS Enterococcus faecalis.
 PN WO9850554-A2.
 PD 12-NOV-1998.
 PF 04-MAY-1998; U08959.
 PR 14-NOV-1997; US-066009.
 PR 06-MAY-1997; US-044031.
 PR 16-MAY-1997; US-046855.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Bailey C, Choi GH, Hromockyj A, Kunsch CA;
 DR WPI; 99-070095/06.
 DR P-PSDB; Y00235.
 DT New isolated Enterococcus faecalis polynucleotides - used to develop
 PT products for the detection of Enterococcus and for use in vaccines
 PT for prevention or attenuation of Enterococcus infection
 PS Claim 1; Page 219; 301pp; English.
 CC The present sequence encodes an antigenic polypeptide fragment
 CC isolated from Enterococcus faecalis. The present invention describes
 CC genes, proteins and antigenic polypeptides isolated from E. faecalis.
 CC The proteins can be used in vaccines for preventing or attenuating an
 CC infection caused by a member of the Enterococcus genus in an animal.
 CC They can also be used for detecting Enterococcus antibodies in a sample.
 CC The nucleotide sequences can be used for detecting Enterococcus nucleic
 CC acids. Products from the present invention can also be used for
 CC screening compounds to identify agonists and antagonists of E. faecalis
 CC protein activity.
 SQ Sequence 1883 BP; 666 A; 357 C; 404 G; 456 T;

Query Match 1.2%; Score 17; DB 1; Length 1883;
 Best Local Similarity 100.0%; Pred. No. 25;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 416 cagtgctgtaaaatgcc 432
 |||||
 DB 1005 CAGTTGCTAAATAATGCC 1021

RESULT 14

X20306
 DT 20306 standard; DNA; 2019 BP.
 AC X20306;
 DT 04-MAY-1999 (first entry)
 DE Borrelia burgdorferi polynucleotide sequence #59.
 KW Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;
 KW epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;
 KW infection; diagnosis; characterisation; detection; ds.
 OS Borrelia burgdorferi.
 PN WO9858943-A1.
 PD 30-DEC-1998.
 PF 18-JUN-1998; U12764.
 PR 03-SEP-1997; US-057483.
 PR 20-JUN-1997; US-050359.
 PR 22-JUL-1997; US-053344.
 PR 22-JUL-1997; US-053377.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI (MEDI-) MEDIMMUNE INC.
 PI Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO,
 PI White OR.
 DR WPI; 99-081217/07.

DT New isolated Borrelia burgdorferi nucleic acids - used to develop
 PT products for the detection, diagnosis, characterisation, prevention
 PT and therapy of infections, particularly Lyme disease
 PS Claim 1; Page 1040-1041; 1128pp; English.
 CC X20248 to X20402 represent polynucleotide sequences isolated from
 CC Borrelia burgdorferi (Bb). Products derived from Bb can be used for
 CC the detection, diagnosis, characterisation, prevention and therapy of
 CC Bb infections, e.g. Lyme disease. They can also be used for the
 CC production of biosynthetic products, e.g. enzymes. Borrelia belongs

CC to a family of motile, spiral-shaped bacteria called Spirochetes.
 CC Spirochetes are pathogenic in humans and Borrelia causes epidemic and
 CC endemic relapsing fever, and Lyme borreliosis, more commonly known as
 CC Lyme disease.
 SQ Sequence 2019 BP; 611 A; 312 C; 298 G; 798 T;

Query Match 1.2%; Score 17; DB 1; Length 2019;
 Best Local Similarity 100.0%; Pred. No. 25;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 128 tgatgatcttttaatgc 144
 |||||
 DB 1906 TGATGATCTTTTAATGC 1922

RESULT 15

X28141/c
 DT 28141 standard; DNA; 3800 BP.
 AC X28141;
 DT 11-JUN-1999 (first entry)
 DE CCoAOMT promoter, PtCCoAOMT1.
 DE CCoAOMT promoter; Poplar; caffeoyl-CoA-O-methyltransferase promoter;
 KW PtCCoAOMT1; plant specific promoter; xylem ray cell;
 KW lignin level alteration; ss.
 OS Populus trichocarpa.
 PN WO9909188-A2.
 PD 25-FEB-1999.
 PF 27-JUL-1998; E04988.
 PR 13-AUG-1997; EP-202507.
 PA (VIAA-) VIAAMS INTERUNIVERSITAIR INST BIOTECHNOG.
 PI Boerjan W, Chen C, Meyermans H, Van Montagu M;
 DR WPI; 99-190169/16.
 PT Poplar caffeoyl-CoA-O-methyltransferase (CCoAOMT) promoter - to
 PT provide transgenic plants having altered lignin level and/or
 PT composition
 PS Claim 1; Fig 1b; 69pp; English.
 CC This sequence represents the Poplar (Populus trichocarpa)
 CC caffeoyl-CoA-O-methyltransferase (CCoAOMT) promoter, PtCCoAOMT1, of the
 CC invention. The promoter is specific for plant or tree living vessel
 CC elements and/or xylem adjacent ray cells. The promoter is used in genetic
 CC modification to provide transgenic plants having altered lignin level
 CC and/or composition, which improves digestibility of forages and pulping
 CC properties of trees.
 SQ Sequence 3800 BP; 1225 A; 668 C; 701 G; 1206 T;

Query Match 1.2%; Score 17; DB 1; Length 3800;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 924 atttgatcaagaactcg 940
 |||||
 DB 2765 ATTTGATCAAGAACTGG 2749

Search completed: September 9, 2000, 22:03:12
 Job time: 2194 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 9, 2000, 21:11:03 ; Search time 51.17 Seconds
(without alignments)
3876.402 Million cell updates/sec

Title: us-09-214-679-1

Perfect score: 1442

Sequence: 1 ccgggaactccatgtgcc.....aatgcaattcatttgatcc 1442

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 243080 seqs, 68777915 residues

Word size : 0

Total number of hits satisfying chosen parameters: 486160

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/5C_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/5D_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/6_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/PCrus_COMB.seq:*
7: /cgn2_6/ptodata/2/ina/backfiles.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	17	1.2	26	1	US-08-039-866-2
2	17	1.2	26	1	US-08-039-866-3
3	16	1.1	75	1	US-07-731-157A-38
4	16	1.1	75	4	US-08-541-780-38
5	16	1.1	109	3	US-08-417-210A-122
6	16	1.1	109	3	US-08-417-210A-145
7	16	1.1	475	3	US-08-967-101-32
8	16	1.1	475	4	US-08-592-541-32
9	16	1.1	1350	1	US-08-665-435A-1
10	16	1.1	1350	1	US-08-665-435A-3
11	16	1.1	1350	3	US-08-843-309-3
12	16	1.1	1353	3	US-08-843-309-1
13	16	1.1	1532	3	US-08-037-816A-15
14	16	1.1	1532	3	US-08-037-816A-27
15	16	1.1	1532	3	US-08-530-146-15
16	16	1.1	1532	3	US-08-530-146-27
17	16	1.1	2019	5	US-08-926-636-1
18	16	1.1	2020	3	US-08-417-210A-135
19	16	1.1	2028	3	US-08-417-210A-138
20	16	1.1	2060	3	US-08-417-210A-141
21	16	1.1	2161	1	US-08-399-646-11
22	16	1.1	2161	2	US-08-607-321-11
23	16	1.1	2161	1	US-08-961-240-11
24	16	1.1	2161	3	US-08-605-501-11
25	16	1.1	2163	1	US-07-731-157A-1
26	16	1.1	2163	4	US-08-541-780-1

27 16 1.1 2193 4 US-08-934-481-1 Sequence 1, Appli
28 16 1.1 2552 3 US-08-448-603A-27 Sequence 27, Appli
29 16 1.1 2552 5 US-09-134-075-27 Sequence 27, Appli
30 16 1.1 2571 1 US-08-254-358-3 Sequence 3, Appli
31 16 1.1 2571 2 US-08-475-391-3 Sequence 3, Appli
32 16 1.1 2571 3 US-08-709-609-3 Sequence 3, Appli
33 16 1.1 2571 6 PCT-US95-07178-3 Sequence 3, Appli
34 16 1.1 2995 2 US-08-592-126-85 Sequence 85, Appli
35 16 1.1 3249 1 US-08-106-493A-1 Sequence 1, Appli
36 16 1.1 3249 1 US-08-429-264-1 Sequence 1, Appli
37 16 1.1 3807 3 US-08-417-210A-78 Sequence 78, Appli
38 16 1.1 4853 2 US-08-832-883-1 Sequence 1, Appli
39 16 1.1 4853 3 US-08-832-877-1 Sequence 1, Appli
40 16 1.1 7797 4 US-08-816-1558-7 Sequence 7, Appli
41 16 1.1 8535 5 US-08-716-351A-1 Sequence 1, Appli
42 16 1.1 8654 2 US-08-920-812-6 Sequence 6, Appli
43 16 1.1 8654 2 US-08-920-827-6 Sequence 6, Appli
44 16 1.1 8654 2 US-08-921-177-6 Sequence 6, Appli
45 16 1.1 8654 2 US-08-362-577C-6 Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-08-039-866-2/c
; Sequence 2, Application US/08039866
; Patent No. 5491079
; GENERAL INFORMATION:
; APPLICANT: KNOL, Jan
; APPLICANT: MARCISET, Olivier
; APPLICANT: MOLLET, Beat
; TITLE OF INVENTION: Integrative gene-expression in
; TITLE OF INVENTION: food-grade microorganisms
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: VOGT & O'DONNELL - STEPHEN M. HARACZ
; STREET: 707 Westchester Avenue
; CITY: WHITE PLAINS
; STATE: NY
; COUNTRY: USA
; ZIP: 10604
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/039,866
; FILING DATE: 30-MAR-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 92105973
; FILING DATE: 07-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: HARACZ, STEPHEN M.
; REGISTRATION NUMBER: 33397
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (914) 328-0055
; TELEFAX: (914) 328-0060
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 26 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-039-866-2

Query Match 1.2%; Score 17; DB 1; Length 26;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 185 ggagaataacatatgaa 201
      |||||
Db 23 GGAGATAACATATGAA 7

RESULT 2
US-08-039-866-3
; Sequence 3, Application US/08039866
; Patent No. 5491079
; GENERAL INFORMATION:
; APPLICANT: KNOL, Jan
; APPLICANT: MARCISSET, Olivier
; APPLICANT: MOLLET, Beat
; TITLE OF INVENTION: Integrative gene-expression in
; TITLE OF INVENTION: food-grade microorganisms
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: VOGT & O'DONNELL - STEPHEN M. HARACZ
; STREET: 707 Westchester Avenue
; CITY: WHITE PLAINS
; STATE: NY
; COUNTRY: USA
; ZIP: 10604
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/039,866
; FILING DATE: 30-MAR-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 92105973
; FILING DATE: 07-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: HARACZ, STEPHEN M.
; REGISTRATION NUMBER: 33397
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (914) 328-0055
; TELEFAX: (914) 328-0060
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 26 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-039-866-3

Query Match 1.2%; Score 17; DB 1; Length 26;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 185 ggagaataacatatgaa 201
      |||||
Db 4 GGAGATAACATATGAA 20

RESULT 3
US-07-731-157A-38
; Sequence 38, Application US/07731157A
; Patent No. 5457032
; GENERAL INFORMATION:
; APPLICANT: Quax, Wilhelmus J.
; APPLICANT: Misset, Onno
; APPLICANT: Van der Laan, Jan M.
; APPLICANT: Lenting, Herman B.M.
; TITLE OF INVENTION: Mutated beta-lactam acylase genes
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: VOGT & O'DONNELL - STEPHEN M. HARACZ
; STREET: 707 Westchester Avenue
; CITY: WHITE PLAINS
; STATE: NY
; COUNTRY: USA
; ZIP: 10604
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/039,866
; FILING DATE: 30-MAR-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 92105973
; FILING DATE: 07-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: HARACZ, STEPHEN M.
; REGISTRATION NUMBER: 33397
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (914) 328-0055
; TELEFAX: (914) 328-0060
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 26 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-039-866-3

Query Match 1.1%; Score 16; DB 1; Length 75;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 838 tatttggtatgcccat 853
      |||||
Db 33 TATTGGTATGCCCAT 48

RESULT 4
US-08-541-780-38
; Sequence 38, Application US/08541780
; Patent No. 5935831
; GENERAL INFORMATION:
; APPLICANT: Quax, Wilhelmus J.
; APPLICANT: Misset, Onno
; APPLICANT: Van der Laan, Jan M.
; APPLICANT: Lenting, Herman B.M.
; TITLE OF INVENTION: Mutated beta-lactam acylase genes
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: COOLEY GODWARD CASTRO HUDDLESON & TATUM
; STREET: FIVE PALO ALTO SQUARE, 4TH FLOOR
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/541,780
; FILING DATE:
; CLASSIFICATION: 435
```


PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/97/731.157
; FILING DATE:
; APPLICATION NUMBER: EP 90200962
; FILING DATE: 18-APR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: RAE-VENTER PH.D., BARBARA
; REGISTRATION NUMBER: 32,750
; REFERENCE/DOCKET NUMBER: GBRO-027/0005
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-494-7622
; TELEFAX: 415-857-0663
; TELEX: 380816 COOLEY PA
; INFORMATION FOR SEQ ID NO: 38:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 75 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (synthetic)
; HYPOTHETICAL: NO
US-08-541-780-38

Query Match 1.1%; Score 16; DB 4; Length 75;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 838 tattggtatgcccat 853
|||||
DB 33 TATTGTTGATGCCCAT 48

RESULT 5
US-08-417-210A-122
; Sequence 122, Application US/08417210A
; Patent No. 5863542
; GENERAL INFORMATION:
; APPLICANT: PAOLETTI, ENZO
; APPLICANT: TARTAGLIA, JAMES
; APPLICANT: COX, WILLIAM I.
; TITLE OF INVENTION: IMMUNODEFICIENCY RECOMBINANT POXVIRUS
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CURTIS, MORRIS & SAFFORD, P.C.
; STREET: 530 FIFTH AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/417,210A
; FILING DATE: 05-APR-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: KOWALSKI, THOMAS J.
; REGISTRATION NUMBER: 32,147
; REFERENCE/DOCKET NUMBER: 454310-2690
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-840-3333
; INFORMATION FOR SEQ ID NO: 122:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 109 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-417-210A-122

Query Match 1.1%; Score 16; DB 3; Length 109;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 780 atataggaccaggag 795
|||||
DB 53 ATATAGGACCAGGAG 68

RESULT 6
US-08-417-210A-145
; Sequence 145, Application US/08417210A
; Patent No. 5863542
; GENERAL INFORMATION:
; APPLICANT: PAOLETTI, ENZO
; APPLICANT: TARTAGLIA, JAMES
; APPLICANT: COX, WILLIAM I.
; TITLE OF INVENTION: IMMUNODEFICIENCY RECOMBINANT POXVIRUS
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CURTIS, MORRIS & SAFFORD, P.C.
; STREET: 530 FIFTH AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/417,210A
; FILING DATE: 05-APR-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: KOWALSKI, THOMAS J.
; REGISTRATION NUMBER: 32,147
; REFERENCE/DOCKET NUMBER: 454310-2690
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-840-3333
; INFORMATION FOR SEQ ID NO: 145:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 109 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-417-210A-145

Query Match 1.1%; Score 16; DB 3; Length 109;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 780 atataggaccaggag 795
|||||
DB 53 ATATAGGACCAGGAG 68

RESULT 7
US-08-967-101-32
; Sequence 32, Application US/08967101
; Patent No. 5840540
; GENERAL INFORMATION:
; APPLICANT: ST. GEORGE-HYSLOP, PETER H
; APPLICANT: ROMMENS, JOHANNA M
; APPLICANT: FRASER, PAUL E
; TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
; TITLE OF INVENTION: TO ALZHEIMER'S DISEASE
; NUMBER OF SEQUENCES: 183

Tue Sep 12 08:47:28 2000

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: TESTA, HURWITZ & THIBEAULT
STREET: High Street Tower - 125 High Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/967,101
FILING DATE: 10-NOV-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/592,541
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Pitcher, Edmund R.
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 248-7000
TELEFAX: (617) 248-7100
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 475 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-967-101-32

```

Query Match 1.1%; Score 16; DB 3; Length 475;
 Best Local Similarity 100.0%; Pred. No. 44;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 972 aaaaattattgagtat 987
|||||
DB 37 AAAATATTATGAGTAT 52

```

```

RESULT 8
US-08-592-541-32
Sequence 32, Application US/08592541
Patent No. 5986054
GENERAL INFORMATION:
APPLICANT: ST. GEORGE-HYSLOP, PETER H
APPLICANT: ROMMENS, JOHANNA M
APPLICANT: FRASER, PAUL E
TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
TITLE OF INVENTION: TO ALZHEIMER'S DISEASE
NUMBER OF SEQUENCES: 183
CORRESPONDENCE ADDRESS:
ADDRESSEE: TESTA, HURWITZ & THIBEAULT
STREET: High Street Tower - 125 High Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/592,541
FILING DATE:
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Pitcher, Edmund R.
TELECOMMUNICATION INFORMATION:

```

```

TELEPHONE: (617) 248-7000
TELEFAX: (617) 248-7100
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 475 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-592-541-32

```

Query Match 1.1%; Score 16; DB 4; Length 475;
 Best Local Similarity 100.0%; Pred. No. 44;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 972 aaaaattattgagtat 987
|||||
DB 37 AAAATATTATGAGTAT 52

```

```

RESULT 9
US-08-665-435A-1
Sequence 1, Application US/08665435A
Patent No. 5681694
GENERAL INFORMATION:
APPLICANT: Skatrud, Paul
APPLICANT: Peery, Robert
APPLICANT: Hoskins, JoAnn
APPLICANT: Wu, Chyun-Yeh Earnest
TITLE OF INVENTION: Biosynthetic Gene MurD of Streptococcus
TITLE OF INVENTION: pneumoniae
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Eli Lilly and Company
STREET: Lilly Corporate Center
CITY: Indianapolis
STATE: Indiana
COUNTRY: US
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/665,435A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Webster, Thomas D
REGISTRATION NUMBER: 39,872
REFERENCE/DOCKET NUMBER: X-9900
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-3334
TELEFAX: 317-276-3861
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1350 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1350
US-08-665-435A-1

```

Query Match 1.1%; Score 16; DB 1; Length 1350;
 Best Local Similarity 100.0%; Pred. No. 45;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1117 gctgggcaacatggtc 1132
|||||
Db 409 CTGGGGCAACATGGTC 424

RESULT 10

US-08-665-435A-3
; Sequence 3, Application US/08665435A
; Patent No. 5681694
; GENERAL INFORMATION:
; APPLICANT: Skatrud, Paul
; APPLICANT: Peery, Robert
; APPLICANT: Hoskins, JoAnn
; APPLICANT: Wu, Chyun-Yeh Earnest
; TITLE OF INVENTION: Biosynthetic Gene MurD of Streptococcus
; TITLE OF INVENTION: pneumoniae
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: US
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/665,435A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Webster, Thomas D
; REGISTRATION NUMBER: 39,872
; REFERENCE/DOCKET NUMBER: X-9900
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-3334
; TELEFAX: 317-276-3861
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1350 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-665-435A-3

Query Match 1.1%; Score 16; DB 1; Length 1350;
Best Local Similarity 81.2%; Pred. No. 45;
Matches 13; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1117 gctgggcaacatggtc 1132
|||||
Db 409 GCUGGGCAACAUGGUC 424

RESULT 11

US-08-843-309-3
; Sequence 3, Application US/08843309
; Patent No. 5834270
; GENERAL INFORMATION:
; APPLICANT: Skatrud, Paul
; APPLICANT: Peery, Robert
; APPLICANT: Hoskins, JoAnn
; APPLICANT: Wu, Chyun-Yeh Earnest
; TITLE OF INVENTION: Biosynthetic Gene Mur D of Streptococcus pneumoniae

; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: US
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/843,309
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Webster, Thomas D
; REGISTRATION NUMBER: 39,872
; REFERENCE/DOCKET NUMBER: X-9900
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-3334
; TELEFAX: 317-276-3861
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1350 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-843-309-3

Query Match 1.1%; Score 16; DB 3; Length 1350;
Best Local Similarity 81.2%; Pred. No. 45;
Matches 13; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1117 gctgggcaacatggtc 1132
|||||
Db 409 GCUGGGCAACAUGGUC 424

RESULT 12

US-08-843-309-1
; Sequence 1, Application US/08843309
; Patent No. 5834270
; GENERAL INFORMATION:
; APPLICANT: Skatrud, Paul
; APPLICANT: Peery, Robert
; APPLICANT: Hoskins, JoAnn
; APPLICANT: Wu, Chyun-Yeh Earnest
; TITLE OF INVENTION: Biosynthetic Gene Mur D of Streptococcus pneumoniae
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: US
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/843,309
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:

```

; NAME: Webster, Thomas D
; REGISTRATION NUMBER: 39,872
; REFERENCE/DOCKET NUMBER: X-9500
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-3334
; TELEFAX: 317-276-3861
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1353 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1353
; US-08-843-309-1

Query Match 1.1%; Score 16; DB 3; Length 1353;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1117 gctgggcaacatggtc 1132
|||||
DB 409 GCTGGGCAACATGGTC 424

RESULT 13
US-08-037-816A-15
; Sequence 15, Application US/08037816A
; Patent No. 5869624
; GENERAL INFORMATION:
; APPLICANT: Hasel, Karl W.
; APPLICANT: Maddon, Paul J.
; TITLE OF INVENTION: HIV-1 VACCINES, ANTIBODY COMPOSITIONS RELATED
; TITLE OF INVENTION: THEREOF, AND THERAPEUTIC AND PROPHYLACTIC USES
; TITLE OF INVENTION: THEREOF
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/037,816A
; FILING DATE: 26-MAR-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 41190/JPW/AJM
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 977-9550
; TELEFAX: (212) 664-0525
; TELEX: 422523 COOPU
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1532 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1522
; OTHER INFORMATION:
; US-08-037-816A-27

Query Match 1.1%; Score 16; DB 3; Length 1532;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 780 atattagaccaggag 795
|||||
DB 932 ATATTAGACCAGGAG 947

RESULT 14
US-08-037-816A-27
; Sequence 27, Application US/08037816A
; Patent No. 5869624
; GENERAL INFORMATION:
; APPLICANT: Hasel, Karl W.
; APPLICANT: Maddon, Paul J.
; TITLE OF INVENTION: HIV-1 VACCINES, ANTIBODY COMPOSITIONS RELATED
; TITLE OF INVENTION: THEREOF, AND THERAPEUTIC AND PROPHYLACTIC USES
; TITLE OF INVENTION: THEREOF
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/037,816A
; FILING DATE: 26-MAR-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 41190/JPW/AJM
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 977-9550
; TELEFAX: (212) 664-0525
; TELEX: 422523 COOPU
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1532 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1522
; OTHER INFORMATION:
; US-08-037-816A-27

Query Match 1.1%; Score 16; DB 3; Length 1532;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 780 atattagaccaggag 795
|||||
DB 932 ATATTAGACCAGGAG 947
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 9, 2000, 20:45:57 ; Search time 585.75 Seconds
(without alignments)
10856.120 Million cell updates/sec

Title: US-09-214-679-1
Perfect score: 1442
Sequence: 1 cccgggaactccatgtggcc.....aatgcaatcatttgatcc 1442

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 5247842 seqs, 2204914090 residues

Word size : 0
Total number of hits satisfying chosen parameters: 10495684

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database : EST:*

1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*

44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_est45:*
80: gb_est46:*
81: gb_est47:*
82: em_est35:*
83: em_est36:*
84: em_est37:*
85: gb_est48:*
86: gb_est49:*
87: gb_est50:*
88: gb_est51:*
89: gb_est52:*
90: gb_est53:*
91: gb_est54:*
92: gb_est55:*
93: gb_gss1:*
94: gb_gss2:*
95: gb_gss3:*
96: gb_gss4:*
97: em_gss1:*
98: em_gss2:*
99: em_gss3:*
100: em_gss4:*
101: gb_gss5:*
102: gb_gss6:*
103: gb_gss7:*
104: gb_gss8:*
105: gb_gss9:*
106: em_gss5:*
107: em_gss6:*
108: em_gss7:*
109: em_gss8:*
110: em_gss9:*
111: em_gss10:*
112: em_gss11:*
113: gb_gss10:*
114: gb_gss11:*
115: em_gss12:*
116: gb_gss12:*

117: gb_gss13:*
 118: gb_gss14:*
 119: gb_gss15:*
 120: gb_gss16:*
 121: gb_gss17:*
 122: gb_gss18:*
 123: gb_gss19:*
 124: em_gss13:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
C 1	21	1.5	409	120	B44876 HS-1060-Al-
C 2	21	1.5	454	91	W66727 me26d06.r1
C 3	20	1.4	240	71	AW357770 41461 MAR
C 4	20	1.4	245	71	AW356799 39389 MAR
C 5	20	1.4	261	71	AW326093 18169 MAR
C 6	20	1.4	279	71	AW357182 40066 MAR
C 7	20	1.4	329	87	N74414 za3c02.s1
C 8	20	1.4	382	91	W87700 zh65b11.s1
C 9	20	1.4	420	47	AU006865
C 10	20	1.4	550	103	AQ529126
C 11	20	1.4	572	81	C79644
C 12	20	1.4	595	104	AQ597255
C 13	19	1.3	330	35	AI146143
C 14	19	1.3	360	51	AV193600
C 15	19	1.3	376	81	C68576
C 16	19	1.3	378	51	AV203736
C 17	19	1.3	429	86	M89105
C 18	19	1.3	440	31	AA799513
C 19	19	1.3	456	41	AI555490
C 20	19	1.3	484	120	B36341
C 21	19	1.3	509	86	H76674
C 22	19	1.3	510	93	AQ036697
C 23	19	1.3	527	47	AI995150
C 24	19	1.3	557	72	AW435445
C 25	19	1.3	563	42	AI683139
C 26	19	1.3	592	119	AZ037112
C 27	19	1.3	636	114	AQ803720
C 28	19	1.3	743	63	AW081978
C 29	18	1.2	227	93	AQ079435
C 30	18	1.2	300	80	C34626
C 31	18	1.2	300	80	C37085
C 32	18	1.2	301	35	AI112615
C 33	18	1.2	311	46	AI917291
C 34	18	1.2	339	80	C15144
C 35	18	1.2	387	81	C82966
C 36	18	1.2	387	81	C83822
C 37	18	1.2	391	81	D86807
C 38	18	1.2	396	80	AW738993
C 39	18	1.2	403	86	H65770
C 40	18	1.2	410	27	AA487512
C 41	18	1.2	414	35	AI144901
C 42	18	1.2	422	80	AW735528
C 43	18	1.2	424	80	AW735654
C 44	18	1.2	429	29	AA652860
C 45	18	1.2	438	93	AQ022853

ALIGNMENTS

RESULT 1
 B44876/c
 LOCUS
 DEFINITION
 HS-1060-Al-C06-MR.abi
 sapiens genomic clone Plate-CT 782 Col-11 Row-M, genomic

21-OCT-1997

GSS

DNA

409 bp

HS-1060-Al-C06-MR.abi

CIT Human Genomic Sperm Library C Homo

sapiens genomic clone Plate-CT 782 Col-11 Row-M, genomic

ACCESSION

VERSION B44876.1 GI:2549710

KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 409)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS

Mahairas,G.G., Zackrone,K.D., Smith,T., Tipton,S., Schmidt,S., Traicoff,R., Abajian,C., Blanchard,A., West,A. and Hood,L.E.
 Construction of a Characterized Clone Resource for Genomic Sequencing: Generation and Preliminary Analysis of 20,000 Sequence Tagged Connectors

TITLE

Unpublished (1997)
 On Dec 15, 1999 this sequence version replaced gi:4213071.

JOURNAL

COMMENT Contact: Mahairas GG, Zackrone KD, Hood L

University of Washington

Seattle, WA 98195, USA

Tel: (206) 616-8744

Fax: (206) 685-7301

Email: kzackron@u.washington.edu

Sequence Tagged Connector

Plate: CT 782 row: M column: 11

Class: BAC ends

High quality sequence stop: 409.

FEATURES

Location/Qualifiers

1..409

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="plate=CT 782 Col=11 Row=M"

/clone_lib="CIT Human Genomic Sperm Library C"

/sex="M"

/notes="Organ: sperm; Vector: pBelOBAC11; BAC Clones in E-Coli DH10B"

BASE COUNT 110 a 82 c 90 g 127 t

ORIGIN

Query Match 1..5% Score 21; DB 120; Length 409;
 Best Local Similarity 100.0%; Pred. No. 0.85; Indels 0; Gaps 0;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 197 atgaatgggttggaagaatcc 217

|||||

Db 179 ATGAATGGTTGGGAAGATCC 159

RESULT 2

W66727/c

LOCUS

DEFINITION

W66727

me26d06.r1 Soares mouse embryo NbMel3.5 14.5 Mus musculus CDNA

clone IMAGE:388619 5', mRNA sequence.

W66727

W66727.1 GI:1375662

EST.

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (1996)

On Jan 7, 1998 this sequence version replaced gi:636815.

Contact: Marra M/Mouse EST Project

WashU-HMI Mouse EST Project

Washington University School of Medicine

WashU-HMI Mouse EST Project

Unpublished (1996)

On Jan 7, 1998 this sequence version replaced gi:636815.

Contact: Marra M/Mouse EST Project

WashU-HMI Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:240451

Seq primer: ETPrimer

High quality sequence stop: 341.

Location/Qualifiers

FEATURES

source

1. 454

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="IMAGE:38619"

/clone_lib="Soares mouse embryo NDME13.5 14.5"

/sex="unknown"

/tissue_type="embryo"

/dev_stage="13.5-14.5dpc total fetus"

/lab_host="DH10B"

/note="vector: pT7T3D-Pac (Pharmacia) with a modified

polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA

was primed with a Not I - oligo(dT) primer [5,

TGTTACCAATCTGAAGTGGGCGCGCGGAAATTTTTTTTTTTTTTTTTT

T 3'], on equal amounts of mRNA from 2 13.5dpc and 2

14.5dpc embryos (total RNA provided by Minoru Ko, Wayne

State Univ., from 2 l; double-stranded cDNA was ligated to

Eco RI adaptors (Pharmacia), digested with Not I and

cloned into the Not I and Eco RI sites of the modified

pT7T3 vector. Library went through one round of

normalization, and was constructed by Bento Soares and

M.Fatima Bonaldo."

99 a 117 c 104 g 134 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity 1.5%; Score 21; DB 91; Length 454;

.Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 904 aatcaccaccatcaagtcca 924

Db 49 AATCACCACCATCAAGTCGA 29

RESULT

AW357770

LOCUS

41461 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.

DEFINITION

AW357770

ACCESSION

AW357770.1

VERSION

AW357770.1

KEYWORDS

EST.

SOURCE

Bos taurus.

ORGANISM

Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

Bovidae; Bovinae; Bos.

1 (bases 1 to 240)

Smith,T.P.I., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,

Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A. and

Keele,J.W.

Design and use of four pooled tissue normalized cDNA libraries for

EST discovery in cattle

Unpublished (2000)

On Jan 6, 2000 this sequence version replaced gi:5676822.

Contact: Smith TPL

USDA, ARS, US Meat Animal Research Center

PO Box 166, Clay Center, NE 68933-0166, USA

Tel: 402 762 4366

Fax: 402 762 4390

Email: smithemail.marc.usda.gov

Single pass sequencing. Bases called and trimmed with phred

v0.980904.e. Vector identified by cross_match with the -minscore 20

and -minmatch 12 options.

PCR Primers

FORWARD: AGGAAACAGCTATGACCAT

BACKWARD: GTTTCCCACTCAGCAGC

Plate: 18 row: N column: 4

Seq primer: ATTTAGTGACACTATAG.

FEATURES

source

1. 240

Location/Qualifiers

/organism="Bos taurus"

/db_xref="taxon:9913"

/clone_lib="MARC 3BOV"

/tissue_type="pooled"

/lab_host="DH10B"

/note="Vector: pCMV SPORT6; Site.1: XbaI; Site.2: XhoI;

Library made from pooled tissue from marrow, alveolar

macrophage, ovary, fetal semitendinosus muscle, and fetal

longissimus muscle."

33 a 75 c 59 g 73 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity 1.4%; Score 20; DB 71; Length 240;

.Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 420 tgctaaaaatgcctttctc 439

Db 15 TGCTAAATGCCCTTCTC 34

RESULT

AW356799

LOCUS

39389 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.

DEFINITION

AW356799

ACCESSION

AW356799.1

VERSION

AW356799.1

KEYWORDS

EST.

SOURCE

Bos taurus.

ORGANISM

Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

Bovidae; Bovinae; Bos.

1 (bases 1 to 245)

Smith,T.P.I., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,

Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A. and

Keele,J.W.

Design and use of four pooled tissue normalized cDNA libraries for

EST discovery in cattle

Unpublished (2000)

On Jan 6, 2000 this sequence version replaced gi:6575823.

Contact: Smith TPL

USDA, ARS, US Meat Animal Research Center

PO Box 166, Clay Center, NE 68933-0166, USA

Tel: 402 762 4366

Fax: 402 762 4390

Email: smithemail.marc.usda.gov

Single pass sequencing. Bases called and trimmed with phred

v0.980904.e. Vector identified by cross_match with the -minscore 20

and -minmatch 12 options.

PCR Primers

FORWARD: AGGAAACAGCTATGACCAT

BACKWARD: GTTTCCCACTCAGCAGC

Plate: 15 row: H column: 16

Seq primer: ATTTAGTGACACTATAG.

Location/Qualifiers

1. 245

/organism="Bos taurus"

/db_xref="taxon:9913"

/clone_lib="MARC 3BOV"

/tissue_type="pooled"

/lab_host="DH10B"

/note="Vector: pCMV SPORT6; Site.1: XbaI; Site.2: XhoI;

Library made from pooled tissue from marrow, alveolar

BASE COUNT ORIGIN	47 a 87 c 47 g 64 t		macrophage, ovary, fetal semitendinosus muscle, and fetal longissimus muscle."			
	Query Match 1.4%; Score 20; DB 71; Length 245;		Best Local Similarity 100.0%; Pred. No. 3.1;			
	Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
Qy	420	tgctaaaaatgccctttctc 439				
Db	147	TGCTAAAAATGCCCTTTCTC 166				
RESULT 5						
	AW326093 261 bp mRNA EST 27-JAN-2000					
	LOCUS 18169 MARC IBOV Bos taurus CDNA 5', mRNA sequence.					
	DEFINITION AW326093					
SOURCE	ACCESSION AW326093.1 GI:6762014					
	VERSION					
	KEYWORDS					
	EST					
ORGANISM	Bos taurus.					
	Bos taurus					
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;					
REFERENCE	Bovidae; Bovinae; Bos.					
	1 (bases 1 to 261)					
	Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,					
	Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A. and					
AUTHORS	Keefe,J.W.					
	Design and use of four pooled tissue normalized cDNA libraries for					
	EST discovery in cattle					
	Unpublished (2000)					
JOURNAL	On Jan 6, 2000 this sequence version replaced gi:6675590.					
	COMMENT					
	Contact: Smith TPL					
	USDA, ARS, US Meat Animal Research Center					
PO BOX 166, Clay Center, NE 68933-0166, USA	Tel: 402 762 4366					
	Fax: 402 762 4390					
	Email: smith@email.marc.usda.gov					
	Single pass sequencing. Bases called and trimmed with phred					
v0.980904.e. Vector identified by cross_match with the -minscore 20	and -minmatch 12 options.					
	PCR Primers					
	FORWARD: AGGAACACGCTATGACCAT					
	BACKWARD: GTTTCCTCCAGTCACGACG					
Plate: 5 row: 1 column: 23	Seq primer: ATTTAGGTGACACTATAG.					
	Location/Qualifiers					
	1..261					
	/organism="Bos taurus"					
FEATURES	/db_xref="taxon:9913"					
	/clone_lib="MARC IBOV"					
	/tissue_type="pooled"					
	/lab_host="DH10B"					
Source	/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;					
	Library made from pooled tissue from lymph node, ovary,					
	fat, hypothalamus, and pituitary."					
	44 a 97 c 55 g 65 t					
BASE COUNT	44 a 97 c 55 g 65 t					
	ORIGIN					
	Query Match 1.4%; Score 20; DB 71; Length 261;					
	Best Local Similarity 100.0%; Pred. No. 3.1;					
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;						
Qy	420	tgctaaaaatgccctttctc 439				
Db	128	TGCTAAAAATGCCCTTTCTC 147				
RESULT 6						
BASE COUNT						
	ORIGIN					
	Query Match 1.4%; Score 20; DB 71; Length 279;					
	Best Local Similarity 100.0%; Pred. No. 3.2;					
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;						
Qy	420	tgctaaaaatgccctttctc 439				
Db	145	TGCTAAAAATGCCCTTTCTC 164				
RESULT 7						
BASE COUNT						
	ORIGIN					
	Query Match 1.4%; Score 20; DB 71; Length 279;					
	Best Local Similarity 100.0%; Pred. No. 3.2;					
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;						
Qy	420	tgctaaaaatgccctttctc 439				
Db	145	TGCTAAAAATGCCCTTTCTC 164				
RESULT 7						
BASE COUNT						
	ORIGIN					
	Query Match 1.4%; Score 20; DB 71; Length 279;					
	Best Local Similarity 100.0%; Pred. No. 3.2;					
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;						
Qy	420	tgctaaaaatgccctttctc 439				
Db	145	TGCTAAAAATGCCCTTTCTC 164				
RESULT 7						
BASE COUNT						
	ORIGIN					
	Query Match 1.4%; Score 20; DB 71; Length 279;					
	Best Local Similarity 100.0%; Pred. No. 3.2;					
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;						
Qy	420	tgctaaaaatgccctttctc 439				
Db	145	TGCTAAAAATGCCCTTTCTC 164				
RESULT 7						
BASE COUNT						
	ORIGIN					
	Query Match 1.4%; Score 20; DB 71; Length 279;					
	Best Local Similarity 100.0%; Pred. No. 3.2;					
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;						
Qy	420	tgctaaaaatgccctttctc 439				
Db	145	TGCTAAAAATGCCCTTTCTC 164				
RESULT 7						
BASE COUNT						
	ORIGIN					
	Query Match 1.4%; Score 20; DB 71; Length 279;					
	Best Local Similarity 100.0%; Pred. No. 3.2;					
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;						
Qy	420	tgctaaaaatgccctttctc 439				
Db	145	TGCTAAAAATGCCCTTTCTC 164				
RESULT 7						
BASE COUNT						
	ORIGIN					
	Query Match 1.4%; Score 20; DB 71; Length 279;					
	Best Local Similarity 100.0%; Pred. No. 3.2;					
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;						
Qy	420	tgctaaaaatgccctttctc 439				
Db	145	TGCTAAAAATGCCCTTTCTC 164				
RESULT 7						
BASE COUNT						
	ORIGIN					
	Query Match 1.4%; Score 20; DB 71; Length 279;					
	Best Local Similarity 100.0%; Pred. No. 3.2;					
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;						
Qy	420	tgctaaaaatgccctttctc 439				
Db	145	TGCTAAAAATGCCCTTTCTC 164				
RESULT 7						
BASE COUNT						
	ORIGIN					
	Query Match 1.4%; Score 20; DB 71; Length 279;					
	Best Local Similarity 100.0%; Pred. No. 3.2;					
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;						
Qy	420	tgctaaaaatgccctttctc 439				
Db	145	TGCTAAAAATGCCCTTTCTC 164				
RESULT 7						
BASE COUNT						
	ORIGIN					
	Query Match 1.4%; Score 20; DB 71; Length 279;					
	Best Local Similarity 100.0%; Pred. No. 3.2;					
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;						
Qy	420	tgctaaaaatgccctttctc 439				
Db	145	TGCTAAAAATGCCCTTTCTC 164				
RESULT 7						

Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)
On Dec 20, 1995 this sequence version replaced gi:1133302.
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: ml3 -40 forward
High quality sequence stop: 308.
Location/Qualifiers
1. 329
/organism="Homo sapiens"
/db_xref="GDB:1241179"
/db_xref="taxon:9606"
/clone="IMAGE:296258"
/clone_lib="Soares fetal liver spleen INFLS"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia)
with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
1st strand cDNA was primed with a Pac I - oligo(dT) primer
[5' AACTGGAAGAATTAAATAAGATCTTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Patima Bonaldo."

BASE COUNT 111 a 81 c 73 g 63 t 1 others
ORIGIN

Query Match 1.4%; Score 20; DB 87; Length 329;
Best Local Similarity 100.0%; Pred. No. 3.2;
-Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 386 gctatcaattcggaacagga 405
|||||
Db 30 GCTATCAATTCGGAACAGGA 49

RESULT 8
W87700 382 bp mRNA EST 02-FEB-1997
DEFINITION zh55b11.s1 Soares_fetal_liver_spleen_INFLS_s1 Homo sapiens cDNA
clone IMAGE:416925 3', mRNA sequence.
W87700
W87700.1 GI:1401754
EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 382)
Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,
Chissoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W.,
Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,
Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,
Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,
Trevisan, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.,
and Marra, M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
9704478
On Jan 5, 1998 this sequence version replaced gi:1327905.
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

TITLE
JOURNAL
MEDLINE
COMMENT

Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 620 Std Error: 0.00
Seq primer: mob.REGA+ET.
Location/Qualifiers
1. 382
/organism="Homo sapiens"
/db_xref="GDB:132539s"
/db_xref="taxon:9606"
/clone="IMAGE:416925"
/clone_lib="Soares_fetal_liver_spleen_INFLS_s1"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia)
with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
This is a subtracted version of the original Soares fetal
liver spleen INFLS library. 1st strand cDNA was primed
with a Pac I - oligo(dT) primer [5'
AACTGGAAGAATTAAATAAGATCTTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Patima Bonaldo."

BASE COUNT 119 a 99 c 93 g 68 t 3 others
ORIGIN

Query Match 1.4%; Score 20; DB 91; Length 382;
Best Local Similarity 100.0%; Pred. No. 3.2;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 386 gctatcaattcggaacagga 405
|||||
Db 33 GCTATCAATTCGGAACAGGA 52

RESULT 9
AU006865 420 bp mRNA EST 31-JUL-1998
LOCUS AU006865 Schizosaccharomyces pombe late log phase cDNA
DEFINITION Schizosaccharomyces pombe cDNA clone spc01051, mRNA sequence.
AU006865
ACCESSION AU006865.1 GI:3343323
VERSION
KEYWORDS EST.
SOURCE fission yeast.
ORGANISM Schizosaccharomyces pombe
Eukaryota; Fungi; Ascomycota; Schizosaccharomycetales;
Schizosaccharomycetaceae; Schizosaccharomycetes.
REFERENCE 1 (bases 1 to 420)
AUTHORS Morimyo, M. and Mita, K.
TITILE Identification of expressed sequence tags of Schizosaccharomyces
pombe
JOURNAL Unpublished (1998)
COMMENT On May 14, 1999 this sequence version replaced gi:4827815.
Contact: Mitsuoki Morimyo
Genome Research Group
National Institute of Radiological Sciences
9-1, Anagawa-4-chome, Inage-ku, Chiba, Chiba 263-8555, Japan
Email: morimyo@nirs.go.jp.
Location/Qualifiers
1. 420
/organism="Schizosaccharomyces pombe"
/strain="972"
/db_xref="taxon:4896"
/clone="spc01051"
/clone_lib="Schizosaccharomyces pombe late log phase cDNA"
/sex="h minus"
/note="Vector: M13mp19; The cDNA library of

FEATURES
source

Schizosaccharomyces pombe was prepared by cloning cDNA into the SmaI site of M13mp19 DNA and the direction of DNA sequences was not always from 5' to 3'. The cDNA data of Schizosaccharomyces pombe are available for searching on the World Wide Web. (URL, <http://www.nirs.go.jp>)

BASE COUNT 156 a 62 c 78 g 124 t
ORIGIN

Query Match 1.4%; Score 20; DB 47; Length 420;
Best Local Similarity 100.0%; Pred. No. 3.2;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 960 tggagatgccgaataatt 979
|||||
Db 331 TGGAGATGCCGAAATATT 350

RESULT 10
LOCUS AQ529126 550 bp DNA GSS 18-MAY-1999
DEFINITION RPCI-11-368P15-TV RPCI-11 Homo sapiens genomic clone RPCI-11-368P15, genomic survey sequence.
ACCESSION AQ529126

VERSION AQ529126.1 GI:4841239
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 550)
AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter,J.C.
TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready Map Building

JOURNAL Unpublished (1997)
COMMENT On Sep 10, 1998 this sequence version replaced gi:3555790.
Other_GSSs: RPCI-11-368P15.TJ
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbeetigr.org
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from Research Genet cs (info@resgen.com). BAC end search page: http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html.
Seq primer: T7
Class: BAC ends.

FEATURES Location/Qualifiers
source 1..550
/organism="Homo sapiens"
/db_xref="GDB:7641302"
/db_xref="taxon:9606"
/clone="RPCI-11-368P15"
/clone_lib="RPCI-11"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; RPCI11 Human Male BAC Library"
BASE COUNT 174 a 80 c 109 g 187 t
ORIGIN

Query Match 1.4%; Score 20; DB 103; Length 550;
Best Local Similarity 100.0%; Pred. No. 3.3;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1178 gtttagtaggaataaataac 1197
|||||
Db 384 GTTTAGTAGGAATAACTAAC 403

RESULT 11
C79644

LOCUS C79644 572 bp mRNA EST 26-JUN-1998
DEFINITION J0069G09 3', mRNA sequence.
ACCESSION C79644

VERSION C79644.1 GI:2519974
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 572)
AUTHORS Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Eukaryota; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Ko,M.S.H., Kitchen,J.R., Wang,X., Wang,X., Threat,T.A., Sun,T., Grahovac,M.J., Mason,S., Lim,M.K., Paonessa,P.D., Sauls,A.D. and Doi,H.

TITLE Systematic analyses of genes expressed in 3.5-dpc mouse blastocyst (The ERATO/Doi Project at Wayne State University)
JOURNAL Unpublished (1997)
COMMENT Contact: Hirofumi Doi

Doi Biosymmetry Project, ERATO
Japan Science and Technology Corporation (JST)
WBG Marine East 12F, 2-6 Nakase, Mihama-ku, Chiba 261-71, Japan
Email: hdbioa.jst.go.jp

FEATURES Location/Qualifiers
source 1..572

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="J0069G09"
/clone_lib="Mouse 3.5-dpc blastocyst cDNA"
/tissue_type="blastocyst"
/dev_stage="3.5-dpc"
BASE COUNT 149 a 136 c 113 g 167 t 7 others
ORIGIN

Query Match 1.4%; Score 20; DB 81; Length 572;
Best Local Similarity 100.0%; Pred. No. 3.3;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 936 actggcagcttctctggcca 955
|||||
Db 354 ACTGGCAGCTTCTCTGGCCA 373

RESULT 12

LOCUS AQ597255 595 bp DNA GSS 08-JUN-1999
DEFINITION HS_5300_H1_A11_T7A RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate-876 Col-21 Row-B, genomic survey sequence.
ACCESSION AQ597255

VERSION AQ597255.1 GI:5028467
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 595)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589

COMMENT

Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Resear h Genetics (info@resgen.com). BAC end Web Server: http://www.hbsc.washington.edu
Plate: 876 row: B column: 21
Seq primer: 77
Class: BAC ends
High quality sequence stop: 595.

FEATURES
source

1. .595
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate-876 Col=21 Row=B"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at EcoRI sites"
173 a 96 c 116 g 193 t 17 others
ORIGIN

Query Match 1.4%; Score 20; DB 104; Length 595;
Best Local Similarity 100.0%; Pred. No. 3.3;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1178 gtttagtaggaataactaac 1197
|||||
Db 421 GTTTAGTAGGAATAACTAAC 440

RESULT 13

A1146143
LOCUS
DEFINITION
UI-R-BT0-qh-h-04-0-UI.s1 UI-R-BT0 Rattus norvegicus cDNA clone
UI-R-BT0-qh-h-04-0-UI 3', mRNA sequence.
ACCESSION
A1146143
VERSION
A1146143.1 GI:3667942
KEYWORDS
EST.
SOURCE
Norway rat.
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 330)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
9704477

On Apr 7, 1998 this sequence version replaced gi:3035792.
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
Oligo-dT track not found, Not I site shown in beginning of sequence
is likely internal to the message. cDNA Library Preparation: M.
Patina Ronaldo, Ph.D. Clone distribution: clones will be available
through Research Genetics This clone is also available through the

I.M.A.G.E. Consortium at LNL (info@image.llnl.gov). IMAGE
ID=178816 The following repetitive elements were found in this
cDNA sequence: 55-85, >GC_rich#Low_complexity
Seq primer: M13 Forward
POLYA=No.

FEATURES
source

1. .330
Location/Qualifiers
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-BT0-qh-h-04-0-UI"
/clone_lib="UI-R-BT0"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; This library
(UI-R-BT0) consists of a mixture of individually tagged
normalized libraries constructed from rat hippocampus,
thalamus, mid-brain, medulla, corpus striatum, cerebral
cortex and testis. The tag used to identify the source
tissue is a string of 3-6 nucleotides present between the
Not I site and the oligo-dT track which allows
identification of the library of origin of a clone within
the mixture. This library was then subtracted using a
driver consisting of a mixture of all clones from UI-R-A0,
UI-R-A1, UI-R-E0, UI-R-EL, UI-R-C0, UI-R-C1, UI-R-C2 and
UI-R-C2p."
42 a 128 c 118 g 42 t
ORIGIN

Query Match 1.3%; Score 19; DB 35; Length 330;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1001 ccgctggaggatgcgcgc 1019
|||||
Db 276 CCGCTGGAGGATGCGCGC 294

RESULT 14
A1193600/c

LOCUS
DEFINITION
AV193600 Yuji Kohara unpublished cDNA:Strain N2 hermaphrodite
embryo Caenorhabditis elegans cDNA clone yk619c5 5', mRNA sequence.
ACCESSION
A193600
VERSION
A193600.1 GI:5575752
KEYWORDS
EST.
SOURCE
Caenorhabditis elegans.
ORGANISM
Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE
1 (bases 1 to 360)
Kohara,Y., Shin-I,T., Thierry-Mieg,J., Thierry-Mieg,D., Mitsuiki,H.,
Nishigaki,A., Motohashi,T., Zeng,Q., Watanabe,H., Sugimoto,A.,
Sano,M., Miyata,A., Mitani,Y., Iida,K., Uesugi,H., Sugiyama,Y. and
Nomoto,H.

Expressed genes in C.elegans

Unpublished (1999)
On Jul 9, 1999 this sequence version replaced gi:5434899.

JOURNAL
COMMENT
Contact: Yuji Kohara
Gene Library Lab
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.

FEATURES
source

1. .360
Location/Qualifiers
/organism="Caenorhabditis elegans"
/strain="N2"
/db_xref="taxon:6239"
/clone="yk619c5"

Tue Sep 12 08:47:29 2000

/clone_lib="Yuji Kohara unpublished cdna:Strain N2
hermaphrodite embryo"
/sex="hermaphrodite"
/dev_stage="embryo"
BASE COUNT 129 a 72 c 50 g 109 t
ORIGIN

Query Match 1.3%; Score 19; DB 51; Length 360;
Best Local Similarity 100.0%; Pred. NO. 12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 195 atatgaatggttgaga 213
|||||
DB 112 ATATGAAATGTTGGAAGA 94

RESULT 15
C68576 376 bp mRNA EST 23-SEP-1997
LOCUS C68576 Yuji Kohara unpublished cdna Caenorhabditis elegans cdna
DEFINITION clone_yk305f6 5', mRNA sequence.
C68576
ACCESSION C68576.1 GI:2429932
VERSION EST.
KEYWORDS Caenorhabditis elegans.
SOURCE Caenorhabditis elegans
ORGANISM Caenorhabditis elegans
REFERENCE
- AUTHORS Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 376)
Kohara,Y., Mochizuki,T., Tabara,H., Watanabe,H., Sugimoto,A.,
Sano,M., Miyata,A. and Nishigaki,A.
Expression map of the C.elegans genome
Unpublished (1996)
On Jun 15, 1998 this sequence version replaced gi:3224348.
COMMENT
Contact: Yuji Kohara
Gene Library Lab
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.
Location/Qualifiers
1..376
/organism="Caenorhabditis elegans"
/strain="CB1489 him-8(el489)"
/db_xref="taxon:6239"
/clone="yk305f6"
/clone_lib="Yuji Kohara unpublished cdna"
/sex="hermaphrodite, male"
/tissue_type="whole animal"
/dev_stage="varied"

BASE COUNT 109 a 75 c 84 g 108 t
ORIGIN

Query Match 1.3%; Score 19; DB 81; Length 376;
Best Local Similarity 100.0%; Pred. NO. 12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 192 aacatgaatggttgga 210
|||||
DB 138 AACATGAAATGTTGGA 156

Search completed: September 9, 2000, 21:38:20
Job time: 3143 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 9, 2000, 20:17:02 ; Search time 1336.1 Seconds
(without alignments)
1936.152 Million cell updates/sec

Title: US-09-214-679-1
Perfect score: 1442
Sequence: 1 cccgggaactccatgtggc.....aatgcaattcatttgatcc 1442

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 972840 seqs, 892348106 residues
Total number of hits satisfying chosen parameters: 1945680

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba1.*
- 2: gb_ba2.*
- 3: gb_om.*
- 4: gb_ov.*
- 5: gb_pat.*
- 6: gb_ph.*
- 7: gb_pl1.*
- 8: gb_pl2.*
- 9: gb_pr1.*
- 10: gb_pr2.*
- 11: gb_pr3.*
- 12: gb_ro.*
- 13: gb_sts.*
- 14: gb_sy.*
- 15: gb_un.*
- 16: em_fun.*
- 17: em_hum1.*
- 18: em_hum2.*
- 19: em_in.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_sy.*
- 29: em_un.*
- 30: em_vi.*
- 31: gb_htg1.*
- 32: gb_htg2.*
- 33: gb_in1.*
- 34: gb_in2.*
- 35: em_ba1.*
- 36: em_ba2.*
- 37: em_hum3.*
- 38: em_hum4.*
- 39: gb_pr4.*
- 40: gb_htg3.*
- 41: gb_htg4.*
- 42: gb_htg5.*
- 43: gb_htg6.*

- 44: gb_htg7.*
- 45: em_htg1.*
- 46: em_htg2.*
- 47: em_htg3.*
- 48: em_hum5.*
- 49: gb_pl3.*
- 50: gb_pr5.*
- 51: gb_htg8.*
- 52: gb_htg9.*
- 53: gb_htg10.*
- 54: gb_htg11.*
- 55: gb_htg12.*
- 56: gb_htg13.*
- 57: gb_htg14.*
- 58: gb_in3.*
- 59: gb_htg15.*
- 60: gb_htg16.*
- 61: gb_htg17.*
- 62: em_htg4.*
- 63: em_htg5.*
- 64: em_htg6.*
- 65: em_htg7.*
- 66: em_hum6.*
- 67: gb_htg18.*
- 68: gb_htg19.*
- 69: gb_htg20.*
- 70: gb_htg21.*
- 71: gb_htg22.*
- 72: gb_htg23.*
- 73: gb_htg24.*
- 74: gb_htg25.*
- 75: gb_htg26.*
- 76: gb_htg27.*
- 77: gb_htg28.*
- 78: gb_htg29.*
- 79: gb_htg30.*
- 80: gb_htg31.*
- 81: gb_v11.*
- 82: gb_v12.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query %			DB	ID	Description
	Score	Match	Length			
1	1442	100.0	1442	5	A69475	A69475 Sequence 1
2	1442	100.0	1442	5	A72152	A72152 Sequence 1
C 3	45.8	3.2	167000	1	AP000059	AP000059 Aeropyrum
C 4	42.8	3.0	35346	8	SPAC869	AL132779 S.pombe c
C 5	41.2	2.9	307120	1	CNSPAX03	AJ248285 Pyrococcus
C 6	41.2	2.9	349061	2	NMA222491	AL162753 Neisseria
C 7	39.6	2.7	92918	43	AC020050	AC020050 Drosophil
C 8	39.6	2.7	123296	51	AC008323	AC008323 Drosophil
C 9	39.6	2.7	130583	41	AC007420	AC007420 Drosophil
C 10	39.6	2.7	304383	34	AE003579	AE003579 Drosophil
C 11	39.4	2.7	118593	10	HS288L1	282196 Human DNA s
C 12	39.2	2.7	10619	2	AE002546	AE002546 Neisseria
C 13	39.2	2.7	133144	42	AC015424	AC015424 Drosophil
C 14	39.2	2.7	133190	51	AC010003	AC010003 Drosophil
C 15	39.2	2.7	154381	51	AC009369	AC009369 Drosophil
C 16	39.2	2.7	271237	34	AE003521	AE003521 Drosophil
C 17	37.6	2.6	503	5	I33795	I33795 Sequence 15
C 18	37.6	2.6	3900	5	I33788	I33788 Sequence 6
C 19	37.6	2.6	3902	33	DMU31226	U31226 Drosophila
C 20	36.8	2.6	29692	42	AC013145	AC013145 Drosophil
C 21	36.8	2.6	102619	33	DMBR37P7	AL050231 Drosophil
C 22	36.8	2.6	314661	34	AE003417	AE003417 Drosophil
C 23	36.6	2.5	39228	1	MLCB1788	AL008609 Mycobacte
C 24	36.6	2.5	167306	32	AL133507	AL133507 Homo sapi


```
Db 1141 ATACACCGTGGCGGATGCTGAACAAACAAACCTGTTAGTTAGTAGGAATAACTAACCGG 1200
Qy 1201 tgaacattaccggatgtagacgcgggtaagtgtgaagtcaacaacaatcgctatttttaa 1260
Db 1201 tgaacattaccggatgtagatcggggtaagtgtgaagtcaacaacaatcgctatttttaa 1260
Qy 1261 cagctaaagcagggtgcataatgaggccagatcacaccatcaatattggtttactactcc 1320
Db 1261 CAGCTAAAGCAGGTGCATATGGGCGCAGATACACCCATCAATATTGGTTTACTTACTCC 1320
Qy 1321 ttcaacgagtagcggcgacacaaagttgtcacaaatggcgggagcaacccagctatt 1380
Db 1321 ttcaacgagtagcggcgacacaaagttgtcacaaatggcgggagcaacccagctatt 1380
Qy 1381 gcccgaattatacaaatggcgcatcaacgcagaccactcaatgcgaattcattggat 1440
Db 1381 GCCGAATTATCAAAATGGCGCATCAACGCCAGACCACCTCAATGCAATTCATTGGAT 1440
Qy 1441 cc 1442
Db 1441 CC 1442

RESULT 2
A72152 LOCUS A72152 1442 bp DNA PAT 11-MAY-1999
DEFINITION Sequence 1 from Patent WO9801568.
ACCESSION A72152
VERSION A72152.1 GI:4808107
KEYWORDS
SOURCE Klebsiella oxytoca.
ORGANISM Klebsiella oxytoca
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Klebsiella.
REFERENCE 1 (bases 1 to 1442)
AUTHORS Brieden, W., Naughton, A., Robins, K., Shaw, N., Tinschert, A. and
Zimmermann, T.
TITLE METHOD OF PREPARING (S) - OR (R) -3,3,3-TRIFLUORO-2-HYDROXY-2-
METHYLPROPIONIC ACID
JOURNAL Patent: WO 9801568-A 15-JAN-1998;
LONZA AG (CH)
FEATURES
source
1..1442 Location/Qualifiers
/organism="Klebsiella oxytoca"
/strain="PRS1"
/isolate="PRS1"
/db_xref="taxon:571"
197..1183
/codon_start=1
/transl_table=11
/product="AMIDASE"
/protein_id="CAB42722.1"
/db_xref="GI:4808108"
/translation="MKWLEESIMAKRGVAGRKPVTHLTEMQKEPHYITGPYSTPV
LTIEPDRILVDTRDAFEAGINSEODIPSQLKMPFLNPQNGPINVNAEKGDLVLA
YTESMLPRGVGIGTCAMTPHEGGLTGDTLAMLNDPLPEKVRMIKIDSEKYSKRHT
LPYKPHIGTSLVSEIDINSILTPDNHGNMDVPDIPGSIYILPVRAPEGRLFTGDA
HACODGRTICATVEFASTITIKVDLIKMWLSWPRMENAINISGARPLEDATRI
AYRDLIWLVEDFGEQWDAYMLLSQCKVRLGNMVDPKYTVGAMLKNLLV"
BASE COUNT 385 a 350 c 370 g 337 t
ORIGIN

Query Match 100.0%; Score 1442; DB 5; Length 1442;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1442; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cccgggaactccatgtggccgtgatcctgtgcagcagatattgcgatgccagcggg 60
Db 1 CCGGGNACTCCATGTGGCGGTGATCCTGTGTCGACGAGATATTGCGATGATCCAGCGG 60
Qy 61 ccgcacagcgtgtggttaataagggcctggttagaanaacgtgaaccaaac 120
Db 1141 ATACACCGTGGCGGATGCTGAACAAACAAACCTGTTAGTTAGTAGGAATAACTAACCGG 1200
```

```
Db 61 CCGCACAGCGCTGTGCGGTAATGGATAAAGGCTGGTGTAGAAACAGCTGACCCACAAC 120
Qy 121 agctctctgatgatcttttaataatgcgtcatctgctctgtaactaaacgcgtataaatt 180
Db 121 AGCTCTCTGATGATCTTTTAATGCGTGCATCTGCTCTGTAACCTAAACGCTATAAATT 180
Qy 181 acgtggagaataacatatgaaatggttggaagaatccattatggccaaacggggtgttg 240
Db 181 ACGTGGAGAATAACATATGAAATGGTTGGAAGAACCATATTGGCCAAACGCGGTGTGG 240
Qy 241 tcccgggcgtaaacggtaaacatcacctgacggaagaataatgcaaaaagagtttcatta 300
Db 241 TGCCGGCGCTAAACCGGTAAACGCATCACCTGACGGGAAGAAATGCAAAAAGAGTTTCATTA 300
Qy 301 caccattggcccttattccacaccgcctctgaccatcgaacccggtagccggtgacgcgatttgt 360
Db 301 CACCATTGGCCCTTATTCCACACCCGCTCTGACCATCGAACCAGCGGTGACCGGATTATTGT 360
Qy 361 cgactcgcagatgcttttgaagggtgctatcaattcgggaacaggatatattccgagccagtt 420
Db 361 CGACACTCGAGATGCTTTTGAAGTGCTATCAATTGGAACAGAGATATTCCGAGCCAGTT 420
Qy 421 gctaaaaatgcccttctcaacccacacaaacggaccgatcattggtcaatggcgcgagaa 480
Db 421 GCTAAAAATGCCCTTTCTCAACCCACAAAACGACCGCATGATGGTCAATGGCCGCGAGAA 480
Qy 481 agtgatgtgcgtcgtctctatcgaaatccatggtgccccgcggcgttgatccctacgg 540
Db 481 AGGTGATGTGCTCGCTGCTCTATATCGAATCCATGTTGCCCGCGGGCTTGATCCCTACGG 540
Qy 541 catctgcgcattgattcccgcatctttggcggaactgacgggagccgacctgaacgccaatgct 600
Db 541 CATCTGCGCATGATTCCCGCATTTTGGCGACTTGACGGGACCGAGCTGACGCCCATGCT 600
Qy 601 caatgctccgtgccagaaaagggtgcgatgatttaaacacacagtgaaagtcactg 660
Db 601 CAATGATCCGCTGCCAGAAAAGGTGCGCATGATTAAACTCGACAGTGAAGAGTCTACTG 660
Qy 661 gagcaaacgccatcacgcttccctataaacccccatttggcaccttgagcgtatcgccaga 720
Db 661 GAGCAAACGCCATACGCTTCCCTATAAACCCCATATTGGCACCTTGAGCGTATCGCCAGA 720
Qy 721 aattgactcaatcaattcactgacgcagacaaatcacgagcggggaatatggatgtgcgga 780
Db 721 AATTGACTCAATCAATTCACTGACGCGCAGACAATCACGGCGGAATATGATGTGCGCGA 780
Qy 781 tataggaccaggagattaccctatctgcggtacgtgcgctggagcgccgctgtttat 840
Db 781 TATAGGACCAGGGAGATTACCTTATCTGCCGGTACGTGGCGCCTGGAGGCGCGCTGTTAT 840
Qy 841 tggatgcccattgctgttcagggtgatggtgagatttgcgggacgcagtagagttgc 900
Db 841 TGGTGTATGCCATGCTGTGTGACGGTGATGGTGAGATTTCGGGGACCGCAGTAGAGTTGC 900
Qy 901 ctcaatcaacacatcaaaagtgcatttgcataagaactggcagctttctctggccacgaat 960
Db 901 CTCAATCACCACCACCATCAAAAGTCGATTGTATCAAGAACTGGCAGCTTTCCTGGCCACGAAT 960
Qy 961 ggagaatgcggaaaaatattatgatgattggcagtcgacgtccgctggagagatcgacgcg 1020
Db 961 GGAGAAATGCCGAAAATATTATGAGTATTGGCAGTCCACGCTCCGCTGGAGATCGCAGCG 1020
Qy 1021 aattgcatacgcgacttaatttacttgctggttagaagactttggcttgcaacaatggga 1080
Db 1021 AATTGCATATCGGACTTTAATTTACTTGGCTGGTAGAAGACTTTGGCTTCGAACAATGGGA 1080
Qy 1081 tgcctacatgcttctgagtcgaatgcggcaaaagtcggcgtgggcaacatgctgcaccccaa 1140
Db 1081 TGCCCTACATGCTTCTGAGTCAATGCGCAAAAGTGGCGTGGGCAACATGGTCCACCCCAA 1140
Qy 1141 atacacgctgtgcgcgatgctgaacaaaaacctgttagtttagtagaataaataaccgg 1200
Db 1141 ATACACCGTGGCGGATGCTGAACAAACAAACCTGTTAGTTAGTAGGAATAACTAACCGG 1200
```

```
QY 1201 tgaacattaccggatgtagatcggggtaatagtgtgtaagttcaacaatcgctatttttaa 1260
|||||
Db 1201 TGAACATTACC GGATG TAGATCGGGTAAATGTGTAAGTTCAACAATCGCTATTTTAA 1260
|||||
QY 1261 cagctaaagcaggtgcataatggggccagacatacccccatcaattgtttactttactcc 1320
|||||
Db 1261 CAGCTAAAGCAGGTGCATATGGGGCCAGATACACCCATCAATATGGTTTACTTTACTCC 1320
|||||
QY 1321 ttacgcaggatgcagcgccagcaagagttgtcacataggcgccagagcaaccagggctatt 1380
|||||
Db 1321 TTACAGCGAGTGCAGCGCCACAGAGTTGTCAACAATGGCGGAGCAACCCAGCGCTATT 1380
|||||
QY 1381 gccgaataatcaaaatggcgccatcaacagcgagacacactcaatgaattcatttgat 1440
|||||
Db 1381 GCCGAATAATCAAAATGGCGCATCAACGGCAGACCACTCAATGCAATTCATTGGAT 1440
|||||
QY 1441 cc 1442
Db 1441 CC 1442

RESULT 3
LOCUS AP000059/c 167000 bp DNA BCT 06-APR-2000
DEFINITION Aeropyrum pernix genomic DNA, section 2/7.
ACCESSION AP000059 BA000002
VERSION AP000059.1 GI:5103911
KEYWORDS
SOURCE Aeropyrum pernix (strain:K1) DNA.
ORGANISM
Archaea; Crenarchaeota; Desulfurococcales; Desulfurococcaceae;
Aeropyrum.
1 (sites)
Kawarabayasi,Y., Hino,Y., Horikawa,H., Yamazaki,S., Halkawa,Y.,
Jin-no,K., Takahashi,M., Sekine,M., Baba,S., Ankai,A., Kosugi,H.,
Hosoyama,A., Fukui,S., Nagai,Y., Nishijima,K., Nakazawa,H.,
Takamiya,M., Masuda,S., Funahashi,T., Tanaka,T., Kudoh,Y.,
Yamazaki,J., Kushida,N., Oguchi,A., Aoki,K., Kubota,K.,
Nakamura,Y., Nomura,N., Sako,Y. and Kikuchi,H.
Complete genome sequence of an aerobic hyper-thermophilic
crenarchaeon, Aeropyrum pernix K1
DNA Res. 6 (2), 83-101 (1999)
99310339
2 (bases 1 to 167000)
Tanaka,T., Hino,Y., Kawarabayasi,Y. and Kikuchi,H.
Direct Submission
Submitted (14-DEC-1998) to the DDBJ/EMBL/GenBank databases. Yutaka
Kawarabayasi, National Institute of Technology and Evaluation,
Biotechnology Center; Nishihara 2-49-10, Shibuya-ku, Tokyo
151-0066, Japan (E-mail:kyutaka@kazusa.or.jp, Tel:+81-3-3481-8951,
Fax:+81-3-3481-8424)
Kawarabayasi,Y. is officially affiliated with the National
Institute of Bioscience and Human-Technology, Tsukuba, Ibaraki
305-0046, Japan.
Nomura, N. and Sako, Y. are at Kyoto University, Sakyo-ku, Kyoto
606-8502, Japan
The other authors are at the National Institute of Technology and
Evaluation, 2-49-10 Nishihara, Shibuya, Tokyo 151-0066, Japan.
All the sequence with length 100 codons or more between ATG or GTG
and stop codon are defined as CDS
Homology analysis is performed by Smith-Waterman algorithm against
GenBank and GenPept release 109; EMBL release 56.0; SwissProt
release 36.0; PIR-Protein release 57.0; and OWL release 31.0.
E-mail address for comments and questions: genome@pe.nite.go.jp
Restriction map, ORF organization, sequence alignment and more
information are available at W.W.W. site of Biotechnology Center,
URL: http://www.miln.nite.go.jp/.
Location/Qualifiers
1..167000
/organism="Aeropyrum pernix"
/strain="K1"
/db_xref="taxon:56636"

FEATURES
source
complement(422..595)
/gene="APES009"
complement(422..595)
/gene="APES009"
/note="similar to PIR:H71068 percent identity:29.167 in
48aa."
/codon_start=1
/transl_table=11
/product="57aa long hypothetical protein"
/protein_id="BAA79228.1"
/db_xref="GI:5103912"
/translation="METAIVYHLANKFHPAIPAIIAGVSDYKAKNKVEFVYEBELKDE
IFEIGLELTLEVK"
complement(1152..1490)
/gene="APE0276"
complement(1152..1490)
/gene="APE0276"
/note="similar to PIR:F69324 percent identity:47.115 in
104aa."
/codon_start=1
/transl_table=11
/product="112aa long hypothetical protein"
/protein_id="BAA79229.1"
/db_xref="GI:5103913"
/translation="MRAGFIRVEEVVDRLVSLFEEFVDTGEGAEIVLAFERNAGLLL
MDDRDARNLAKLGLQVMGTGLVIALAKYKGLTSKAKPIIDKLIESGFWISRRTLEEF
LRKLGSRN"
1228..1722
/gene="APE0277"
1228..1722
/gene="APE0277"
/codon_start=1
/transl_table=11
/product="164aa long hypothetical protein"
/protein_id="BAA79230.1"
/db_xref="GI:5103914"
/translation="MSLSIIGFALEVNPLYLANIMPSVDPITCNPSFLARFLASRSI
SSKPAUKSKARTASRSPVTSNSKRDTSLSSTTSSTRIKPAITTPSLGLPATVISA
QTVSGTTRTPKSFSSRRSTLASAMIMLEITRSLSSSSSLLLKLLGLQLQLLL
LATV"
complement(2280..2756)
/gene="APE0278"
complement(2280..2756)
/gene="APE0278"
/codon_start=1
/transl_table=11
/product="158aa long hypothetical protein"
/protein_id="BAA79231.1"
/db_xref="GI:5103915"
/translation="MDKDEKDGELVDTNIIILASVLRHGSYTRQVVIYLDVLGK
AFTSEKALHEIDEHELEAKRKVSVEELRAARILLNVMVEKGTYNAYMEKAGEC
VKDLADAFALALHUSEKYKNVILLTNKNNDYIENCKRAYGVIITYPSDLRVPVN"
complement(3301..3669)
/gene="APE0279"
complement(3301..3669)
/gene="APE0279"
/note="similar to PIR:H69385 percent identity:39.175 in
97aa."
/codon_start=1
/transl_table=11
/product="122aa long hypothetical protein"
/protein_id="BAA79232.1"
/db_xref="GI:5103916"
/translation="MVCISPHSISVYAPRLFLVEVAGLVRYLAPSLIVERVLDAFSS
KVILVDEAFRIAEVETALATGSGADSYILGLAKTLNLPVATSDKVQAQNAKAGIK
SFYILSNNEEELMKYMCCK"
complement(3752..3979)
/gene="APES010"
complement(3752..3979)
/gene="APES010"
/note="similar to PIR:F69385 percent identity:38.462 in
52aa."
/codon_start=1
```

```
/transl_table=11
/product="75aa long hypothetical protein"
/protein_id="BAA79233.1"
/db_xref="GI:5103917"
/translation="MELRVVEIGMSKVIRVRYEKVLKPLPVLNLEDEEVDIIIREN
LAELARRIRRLSOREEPSEILSRERSLA"
4028..4402
/gene="APE0280"
CDS
4028..4402
/gene="APE0280"
/codon_start=1
/transl_table=11
/product="124aa long hypothetical protein"
/protein_id="BAA79234.1"
/db_xref="GI:5103918"
/translation="MIRSPDYHISWCVPFYNAHMOHSLYSAFSGAAGPWGNSPRRAR
DKTYITINLCLINHLFPGLHTCPKSQIEPPPEELLYRLDYLNMVLTLRHNLSGG
MYRVALEGLSIVGIDMEGDLR"
complement(4516..4842)
/gene="APE0281"
complement(4516..4842)
/gene="APE0281"
/codon_start=1
/transl_table=11
/product="108aa long hypothetical protein"
/protein_id="BAA79235.1"
/db_xref="GI:5103919"
/translation="MVTHTPTVTVLSPMPTSSGVNRMWAGTVPAGNNATEHTAVPAISO
SSPDSFFTCASRTSIAASSAVVSALPPTFPATILTSSTKLLYLGLYFSSPHDKIRKAN
KRWAP"
complement(4803..5465)
/gene="APE0282"
complement(4803..5465)
/gene="APE0282"
/codon_start=1
/transl_table=11
/product="220aa long hypothetical protein"
/protein_id="BAA79236.1"
/db_xref="GI:5103920"
/translation="MSSGILFOYRSNICVYFFHFGSSLFASLIASISMDIKGLFIW
SAYARVLPWYFLAGRACAIISFAYSOLAFRMWYILSLSFPCFPASFRLAVSLQGG
LISLSFLSNSSWYSPMFCNPFMLWGSSASFSLFRSRYSALYSACSCSYASSIVSTMY
WSICSRVSWGALSSMCCFFFLRFLSRSMSPAPRPISLLILPLSLPWSPIQRUPCY
PR"
4831..6840
/gene="APE0283"
4831..6840
/gene="APE0283"
/feature="motif=ATP/GTP-binding site motif A (P-loop)"
/codon_start=1
/transl_table=11
/product="669aa long hypothetical protein"
/protein_id="BAA79237.1"
/db_xref="GI:5103921"
/translation="MGDRHGDGRMSRELIGLAGILIRLLLNKKKKQGHIEERAPPO
LTRQIDQYIVDTIDEAYEQLAQYDRAYRELLKREKEADPHIKLQNIKEYHEL
FRKNEMPPWMLTALLNEALKHGVKDRDKTYOYLKASWEYAKEIAQALPAKRYQG
RYTRAYADQMRNPFTSIIEAMIREANRDPKTKYQTMFDLYWNKIPDDIYNRADS
OHHEVTKYLLLEALREARESVEPWHKTPPEILIGIGEGEPAPAGEPIIATYDPR
YPTLAQVKLLTLVGEELREGGVVHRDLYRALRNLLDSPLTNNHWPFLAEWFENLE
KGRHVSPTLQQLKLPVNPQLQRAYRYLAQIEQKSTSEIPEERKAFREVIEWLKN
EIRNPRDFRVIEELVATGMNIDAAKTIGVLDFAKYSGEKSPQRIERDAT
QVLHVKIKNRDGNVDPYHVHTLINIEQHIOQPLLREAYKRTLEHYQTHGPIPR
DLRYLAEERLEERLEKLGILERNRKLSEGDIFRTAMLYRTLLMTPHPSLEDIKR
FYSRLAEKLVKVGESFELPVHGVPAEDVASSFAAIRAKPEERVRVRRREISGEPL
FHAPHIKEGVNSVLAATRSIVSGEKHEEFIRHLERTQVRRLPE"
complement(5360..5668)
/gene="APE0284"
complement(5360..5668)
/gene="APE0284"
/codon_start=1
/transl_table=11
/product="102aa long hypothetical protein"
```

```
/protein_id="BAA79238.1"
/db_xref="GI:5103922"
/translation="MSGTCPPRRVSALPARDPLVLCRGFPRGSCARRALPPIPLFP
HEGPLPGVAGIILLHDAASRPYRCRPESSCNTDQTSVIFSTSDHPCLLLSSPRL
"
Query Match 3.2%; Score 45.8; DB 1; Length 167000;
Best Local Similarity 53.0%; Pred. No. 0.07;
Matches 98; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

QY 756 acggcggaatgagtgccggaatataggaccaggagattaccctattcccggtac 815
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 161824 ACGAGGAAACCTCGATGTGAGGCACTTCCTCCCGAGGCTCAAAATCTACTTCCCGGTG 161765

QY 816 gtgcgcctgagcgccctgtttattggtgatcccatgcttgcaggggtgatggtaga 875
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 161764 TCGTCGAGGGGGCGCTATTCTCGGTGGCGCACGCCACTATGCCAGGGGCGGCGGAGG 161705

QY 876 ttgcgggaccgcagtagagttgttcctcaatcacccaccatcaaaagtcgattgatcaaga 935
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 161704 TGTGCGGCACAGCATAGAGATGGGGCCACCGCTACCTTAAGGTTTCGGGGTCATAAGCG 161645

QY 936 actgg 940
|||||
Db 161644 AGGGG 161640

RESULT 4
SPAC869/c
LOCUS SPAC869 35346 bp DNA PLN 15-DEC-1999
DEFINITION S.pombe chromosome I cosmid c869.
ACCESSION AL132779
VERSION AL132779.2 GI:6594223
KEYWORDS alpha-galactosidase; amidase; amino-acid permease;
flavohemoprotein; formamidase; globin; oxidoreductase;
protein-l-isopartate o-methyltransferase; sodium:solute symporter
family; sulfate permease; sulfate transporter family; Tfl LTR; urea
active transporter.
fission yeast.
SOURCE Schizosaccharomyces pombe
ORGANISM Eukaryota; Fungi; Ascomycota; Schizosaccharomycetales;
Schizosaccharomycetaceae; Schizosaccharomycetes.
REFERENCE 1 (bases 1 to 35346)
AUTHORS Hunt,C., Aves,S., McDougall,R.C., Rajadream,M.A. and Barrell,B.G.
TITLE Direct Submission
JOURNAL Submitted (02-NOV-1999) European Schizosaccharomycetes genome
sequencing project, Sanger Centre, The Wellcome Trust Genome
Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk
and Department of Biological Sciences, University of Exeter, Perry
Road, Exeter EX4 4QG, United Kingdom
On Dec 16, 1999 this sequence version replaced gi:6224591.
COMMENT
Notes:
Details of yeast sequencing at the Sanger Centre are available on
the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/S_pombe/)
During 1995 to 1996 about 66% of S. pombe chromosome 1 was
sequenced by the Sanger Centre. The sequencing of the S. pombe
genome is now being continued with funding from The European
Commission. Fourteen European sequencing laboratories, including
the Sanger Centre, are participating in the project.
Protein coding regions (CDS) have been predicted with the help of
computer analysis using the Genefinder program in PomBase (an ACEDB
database) with additional predictions for the branch-acceptor sites
supplied by the program Sp3splice. CAUTION: It is possible that for
any individual CDS we may have underestimated or overestimated the
number of introns/exons or we may not have chosen the correct
splice donor/acceptor sites.
CDS are numbered using the following system eg SPBC25H2.01c. SP (S.
pombe), B (chromosome 2), c25H2 (cosmid name), .01 (first CDS), c
(complementary strand).
The more significant matches with motifs in the PROSITE database
are also included but some of these may be fortuitous.
The length in codons is given for each CDS.
```

IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.

FEATURES

```

source
    1. .35346
        /organism="Schizosaccharomyces pombe"
        /strain="972h-"
        /db_xref="taxon:4896"
        /chromosome="I"
        /clone="cosmid c869"
        /map="IR"
        complement(1..1610)
        /gene="SPAC869.11"
        1. .1141
            /note="Overlap with c922 S. pombe chromosome 1"
            complement(1..1610)
            /partial
            /gene="SPAC869.11"
            /note="SPAC869.11, len:536, SIMILARITY:Schizosaccharomyces
            pombe, ISP5_SCHPO, sexual differentiation process putative
            amino-acid permease isp5., (543 aa), fasta scores: Opt:
            1812, E():0, (53.9% identity in 492 aa)"
            /codon_start=1
            /label="SPAC869.11"
            /product="amino-acid permease"
            /protein_id="CAB60021.1"
            /db_xref="GI:6224602"
            translation="MEPEYVSFSDKTSILNESKSLKDVKPSLEKSVITPGLVDD
            VEPKKNVRFDDPAKATGDTALKRSLSRHMOMISIGGALGTLYVGGSS
            LADGSPASVILNYSLIGIMPEFIYALGEMAVYPVAGGENTVATRIDPAGFAVSW
            NYFINVFPELTTCAITPRTDINSAMISILVLIIVNLCVRAVGEVFTIL
            STKVAVTFPIILAILINGGVPTDHRGYIGGSLIHKHPRHGKFGCQSVFTAAS
            FSGTEVIGLAAAEVDNPKALPHAVQVFWRIALFYVVSILIGLLISPDDPNLMNGS
            STVSFVFILAEKANIIGLPSVFNAILISVYTSNSTYTAGRTLHGMANLKOQPSF
            FKYIDRLGRILLAMVLLFFGFAYINEADKNGNDVSDTVFNWLLSLGSLNFTWGS
            ICLCHIIFRLAPKQGHSLKELGFVSPDWGSCIGLGFNFILCLMAQFYVSLFPIGGK
            PNANDFFOGYLAAPVLAFFIGYKIYDRS"
            complement(3..1394)
            /gene="SPAC869.11"
            /note="Match to PF00324 aa_permeases, Amino acid permease
            Score 820.26"
            3705..5363
            /gene="SPAC869.10c"
            3705..5363
            /gene="SPAC869.10c"
            /note="SPAC869.10c, len:552, SIMILARITY:Emericella
            nidulans, PUTX_EMENI, proline-specific permease, (550 aa),
            fasta scores: Opt: 1968, E():0, (53.3% identity in 520
            aa)"
            /codon_start=1
            /label="SPAC869.10c"
            /product="amino-acid permease"
            /protein_id="CAB60020.1"
            /db_xref="GI:6224601"
            translation="MDEKALQYEEESKRAEETDIELVSVGIDVEKKYGETKRALKS
            RVOLTAIGCGICGTGLFVSGSALSESQSPASFLSYVIMSFVITVMNALGEMCTYLP
            LSGASPTIERYVDASLAFAAGWNWYAVFLVASEVTAASIVIEYWTYAVPTGFRY
            ALLFLVAVLNSFFKWKGEETFEFALIKVIALVGLIIVGVIFVFGTPKHDLGFRY
            WKHGLAFREYIYKSGRVGFWSAVKSQFAPILAPELVIFSAGETEARPNRIKAT
            SRFIYRLIEFYIGSLTIGVITSSKDPRLNLIAISGASGAASAPVIGIQNAIEPVLN
            HTINAVILTSAGSNFSFLFAGRSIYSLAKEHQAPKIFKYNCRNGVPVIVSAVTPLF
            ACIALFASAAVNWFCNLSITISGLAWICVAVLQVQFRKAMILANNMETRYTKT
            PQPYATYLTLLTALTNGTFTVFGHTFAGNFIAAYITPLFLVLYVAHLKWSR
            NWSFGRIEIDVTYGVAAEALEQMPYAPVPRNIEIKWFIA"
            3816..5228
            /gene="SPAC869.10c"
            /note="Match to PF00324 aa_permeases, Amino acid permease
            Score 718.22"
            complement(6788..7138)
            /note="Tf1 LTR"
            complement(7524..7874)
            /gene="SPAC869.09"
            complement(7524..7874)
            /gene="SPAC869.09"
            len:116, SIMILARITY:Schizosaccharomyces
            pombe, YAA8_SCHPO, hypothetical 15.4 kd protein c22g7.11c
            in chromosome I., (140 aa), fasta scores: Opt: 192,
            E():2.5e-06, (39.6% identity in 139 aa)"
            /codon_start=1
            /label="SPAC869.09"
            /product="hypothetical protein"
            /protein_id="CAB60019.1"
            /db_xref="GI:6224600"
            translation="MPDPAHIIAGHKAALSNPHVSEAEKERARKYLKHGSSSHYTTG
            TTRGOKADADAGELREEGFTGNQFEDNESQAKNLGNVGGYKAAMNPKVSKGRR
            HAKELLEVDDESK"
            complement(8756..9448)
            /gene="SPAC869.08"
            complement(8756..9448)
            /note="SPAC869.08, len:230, SIMILARITY:Triticum aestivum,
            PMT_WHEAT, protein-l-isoaspartate o-methyltransferase,
            (230 aa), fasta scores: Opt: 602, E():2.2e-32, (48.5%
            identity in 200 aa)"
            /codon_start=1
            /label="SPAC869.08"
            /product="putative protein-l-isoaspartate
            o-methyltransferase"
            /protein_id="CAB60018.1"
            /db_xref="GI:6224599"
            translation="MEWSENLSSNALVQHVLESKELTNORAIKAMNATSRGFCPLS
            PYMDSQSGYGYTISAPHMHATALEQLEVPLOPGCSALDIGSGSYLVAAMRWAP
            NGTVKGIEHPQLVETSKNKLADINHDEVLMYKEKRLQINVGDMGTSEDEKFD
            AIVHGASASLEPKQLVDQLKSPGKILIPITGTSQNIYLIEKNEQGIKSKRTLPVRYV
            PLTDSPPDSSDY"
            complement(8783..9427)
            /gene="SPAC869.08"
            /note="Match to PF01135 PCMT,
            protein-L-isoaspartate(D-aspartate) O-methyltransferase
            (PCMT) Score 263.14"
            10265..11575
            /gene="SPAC869.07c"
            10265..11575
            /gene="SPAC869.07c"
            /note="SPAC869.07c, len:436, SIMILARITY:Mortierella
            vinacea., Q02402, alpha-galactosidase, (417 aa), fasta
            scores: Opt: 1443, E():0, (52.3% identity in 407 aa)"
            /codon_start=1
            /label="SPAC869.07c"
            /product="putative alpha-galactosidase"
            /protein_id="CAB60017.1"
            /db_xref="GI:6224598"
            translation="MISIFLNCFFLVFLFTSDVHGSYNGLGKPKQMGWNSWKYA
            CDDIESIILNNAKAKEEGLLDGYEIVVMDDCWSKHENATTTGRLEANPKPPNGIG
            SMARKLHDMGFKFGYSSAGKTYCAGFPGLSNHQIDADTFADGVDYLYKDYCNFEG
            KSGVPLISYERYKMSDALNKTGRPIFYSLCQMGEDFVWNGNTIANSWRIISGDIPT
            FSRKDVPCPETIECFALQGDHCSVMNIIISKASPLSKAGMNSGNDLSDLEVNGMG
            SFREYKTHFTWAILKSPILGNDVSSMSPKDLIVSNKELISINDIGTNPAAIWK
            KKYDEYIELESGRLNNDVAVLNAASEPLKMGHILSDIFDALGNAEHDWLATDL
            WNNVNLVSKIRANVASHGVQVWFQYKVKNTDKFFSEKHI"
            13058..13669
            /gene="SPAC869.06c"
            13058..13669
            /gene="SPAC869.06c"
            /note="SPAC869.06c, len:203"
            /codon_start=1
            /label="SPAC869.06c"
            /product="hypothetical protein"
            /protein_id="CAB60016.1"
            /db_xref="GI:6224597"
            translation="MLOYSKKKVSLNFPFRLSVKMTFRISDAIFKHRLQSDYQNI
            KSANDYDPTKQNOFWALDARHSVEGIEVYVYKPEFGEKMAEKDRHEQLVK
            EMLYKFSQMKANQSNFIPALDELMSLGKHIDEEOHDIPELEKHLSEESLHMASSF
            ERTKKFVPTHSHPSAPNKPPTETVAGLFAAIDIKLRDMKEKP"
            16383..18905
            /gene="SPAC869.05c"
            16383..18905
            /gene="SPAC869.05c"
            len:116, SIMILARITY:Schizosaccharomyces
            pombe, YAA8_SCHPO, hypothetical 15.4 kd protein c22g7.11c
            in chromosome I., (140 aa), fasta scores: Opt: 192,
            E():2.5e-06, (39.6% identity in 139 aa)"
            /codon_start=1
            /label="SPAC869.05c"
            /product="hypothetical protein"
            /protein_id="CAB60015.1"
            /db_xref="GI:6224596"
            translation="MLOYSKKKVSLNFPFRLSVKMTFRISDAIFKHRLQSDYQNI
            KSANDYDPTKQNOFWALDARHSVEGIEVYVYKPEFGEKMAEKDRHEQLVK
            EMLYKFSQMKANQSNFIPALDELMSLGKHIDEEOHDIPELEKHLSEESLHMASSF
            ERTKKFVPTHSHPSAPNKPPTETVAGLFAAIDIKLRDMKEKP"
            16383..18905
            /gene="SPAC869.04c"
            16383..18905
            /gene="SPAC869.04c"
            len:116, SIMILARITY:Schizosaccharomyces
            pombe, YAA8_SCHPO, hypothetical 15.4 kd protein c22g7.11c
            in chromosome I., (140 aa), fasta scores: Opt: 192,
            E():2.5e-06, (39.6% identity in 139 aa)"
            /codon_start=1
            /label="SPAC869.04c"
            /product="hypothetical protein"
            /protein_id="CAB60014.1"
            /db_xref="GI:6224595"
            translation="MLOYSKKKVSLNFPFRLSVKMTFRISDAIFKHRLQSDYQNI
            KSANDYDPTKQNOFWALDARHSVEGIEVYVYKPEFGEKMAEKDRHEQLVK
            EMLYKFSQMKANQSNFIPALDELMSLGKHIDEEOHDIPELEKHLSEESLHMASSF
            ERTKKFVPTHSHPSAPNKPPTETVAGLFAAIDIKLRDMKEKP"
            16383..18905
            /gene="SPAC869.03c"
            16383..18905
            /gene="SPAC869.03c"
            len:116, SIMILARITY:Schizosaccharomyces
            pombe, YAA8_SCHPO, hypothetical 15.4 kd protein c22g7.11c
            in chromosome I., (140 aa), fasta scores: Opt: 192,
            E():2.5e-06, (39.6% identity in 139 aa)"
            /codon_start=1
            /label="SPAC869.03c"
            /product="hypothetical protein"
            /protein_id="CAB60013.1"
            /db_xref="GI:6224594"
            translation="MLOYSKKKVSLNFPFRLSVKMTFRISDAIFKHRLQSDYQNI
            KSANDYDPTKQNOFWALDARHSVEGIEVYVYKPEFGEKMAEKDRHEQLVK
            EMLYKFSQMKANQSNFIPALDELMSLGKHIDEEOHDIPELEKHLSEESLHMASSF
            ERTKKFVPTHSHPSAPNKPPTETVAGLFAAIDIKLRDMKEKP"
            16383..18905
            /gene="SPAC869.02c"
            16383..18905
            /gene="SPAC869.02c"
            len:116, SIMILARITY:Schizosaccharomyces
            pombe, YAA8_SCHPO, hypothetical 15.4 kd protein c22g7.11c
            in chromosome I., (140 aa), fasta scores: Opt: 192,
            E():2.5e-06, (39.6% identity in 139 aa)"
            /codon_start=1
            /label="SPAC869.02c"
            /product="hypothetical protein"
            /protein_id="CAB60012.1"
            /db_xref="GI:6224593"
            translation="MLOYSKKKVSLNFPFRLSVKMTFRISDAIFKHRLQSDYQNI
            KSANDYDPTKQNOFWALDARHSVEGIEVYVYKPEFGEKMAEKDRHEQLVK
            EMLYKFSQMKANQSNFIPALDELMSLGKHIDEEOHDIPELEKHLSEESLHMASSF
            ERTKKFVPTHSHPSAPNKPPTETVAGLFAAIDIKLRDMKEKP"
            16383..18905
            /gene="SPAC869.01c"
            16383..18905
            /gene="SPAC869.01c"
            len:116, SIMILARITY:Schizosaccharomyces
            pombe, YAA8_SCHPO, hypothetical 15.4 kd protein c22g7.11c
            in chromosome I., (140 aa), fasta scores: Opt: 192,
            E():2.5e-06, (39.6% identity in 139 aa)"
            /codon_start=1
            /label="SPAC869.01c"
            /product="hypothetical protein"
            /protein_id="CAB60011.1"
            /db_xref="GI:6224592"
            translation="MLOYSKKKVSLNFPFRLSVKMTFRISDAIFKHRLQSDYQNI
            KSANDYDPTKQNOFWALDARHSVEGIEVYVYKPEFGEKMAEKDRHEQLVK
            EMLYKFSQMKANQSNFIPALDELMSLGKHIDEEOHDIPELEKHLSEESLHMASSF
            ERTKKFVPTHSHPSAPNKPPTETVAGLFAAIDIKLRDMKEKP"
            16383..18905
            /gene="SPAC869.00c"
            16383..18905
            /gene="SPAC869.00c"
            len:116, SIMILARITY:Schizosaccharomyces
            pombe, YAA8_SCHPO, hypothetical 15.4 kd protein c22g7.11c
            in chromosome I., (140 aa), fasta scores: Opt: 192,
            E():2.5e-06, (39.6% identity in 139 aa)"
            /codon_start=1
            /label="SPAC869.00c"
            /product="hypothetical protein"
            /protein_id="CAB60010.1"
            /db_xref="GI:6224591"
            translation="MLOYSKKKVSLNFPFRLSVKMTFRISDAIFKHRLQSDYQNI
            KSANDYDPTKQNOFWALDARHSVEGIEVYVYKPEFGEKMAEKDRHEQLVK
            EMLYKFSQMKANQSNFIPALDELMSLGKHIDEEOHDIPELEKHLSEESLHMASSF
            ERTKKFVPTHSHPSAPNKPPTETVAGLFAAIDIKLRDMKEKP"
            16383..18905
            /gene="SPAC869.00c"
            16383..18905
            /gene="SPAC869.00c"
            len:116, SIMILARITY:Schizosaccharomyces
            pombe, YAA8_SCHPO, hypothetical 15.4 kd protein c22g7.11c
            in chromosome I., (140 aa), fasta scores: Opt: 192,
            E():2.5e-06, (39.6% identity in 139 aa)"
            /codon_start=1
            /label="SPAC869.00c"
            /product="hypothetical protein"
            /protein_id="CAB60009.1"
            /db_xref="GI:6224590"
            translation="MLOYSKKKVSLNFPFRLSVKMTFRISDAIFKHRLQSDYQNI
            KSANDYDPTKQNOFWALDARHSVEGIEVYVYKPEFGEKMAEKDRHEQLVK
            EMLYKFSQMKANQSNFIPALDELMSLGKHIDEEOHDIPELEKHLSEESLHMASSF
            ERTKKFVPTHSHPSAPNKPPTETVAGLFAAIDIKLRDMKEKP"
            16383..18905
            /gene="SPAC869.00c"
            16383..18905
            /gene="SPAC869.00c"
            len:116, SIMILARITY:Schizosaccharomyces
            pombe, YAA8_SCHPO, hypothetical 15.4 kd protein c22g7.11c
            in chromosome I., (140 aa), fasta scores: Opt: 192,
            E():2.5e-06, (39.6% identity in 139 aa)"
            /codon_start=1
            /label="SPAC869.00c"
            /product="hypothetical protein"
            /protein_id="CAB60008.1"
            /db_xref="GI:6224589"
            translation="MLOYSKKKVSLNFPFRLSVKMTFRISDAIFKHRLQSDYQNI
            KSANDYDPTKQNOFWALDARHSVEGIEVYVYKPEFGEKMAEKDRHEQLVK
            EMLYKFSQMKANQSNFIPALDELMSLGKHIDEEOHDIPELEKHLSEESLHMASSF
            ERTKKFVPTHSHPSAPNKPPTETVAGLFAAIDIKLRDMKEKP"
            16383..18905
            /gene="SPAC869.00c"
            16383..18905
            /gene="SPAC869.00c"
            len:116, SIMILARITY:Schizosaccharomyces
            pombe, YAA8_SCHPO, hypothetical 15.4 kd protein c22g7.11c
            in chromosome I., (140 aa), fasta scores: Opt: 192,
            E():2.5e-06, (39.6% identity in 139 aa)"
            /codon_start=1
            /label="SPAC869.00c"
            /product="hypothetical protein"
            /protein_id="CAB60007.1"
            /db_xref="GI:6224588"
            translation="MLOYSKKKVSLNFPFRLSVKMTFRISDAIFKHRLQSDYQNI
            KSANDYDPTKQNOFWALDARHSVEGIEVYVYKPEFGEKMAEKDRHEQLVK
            EMLYKFSQMKANQSNFIPALDELMSLGKHIDEEOHDIPELEKHLSEESLHMASSF
            ERTKKFVPTHSHPSAPNKPPTETVAGLFAAIDIKLRDMKEKP"
            16383..18905
            /gene="SPAC869.00c"
            16383..18905
            /gene="SPAC869.00c"
            len:116, SIMILARITY:Schizosaccharomyces
            pombe, YAA8_SCHPO, hypothetical 15.4 kd protein c22g7.11c
            in chromosome I., (140 aa), fasta scores: Opt: 192,
            E():2.5e-06, (39.6% identity in 139 aa)"
            /codon_start=1
            /label="SPAC869.00c"
            /product="hypothetical protein"
            /protein_id="CAB60006.1"
            /db_xref="GI:6224587"
            translation="MLOYSKKKVSLNFPFRLSVKMTFRISDAIFKHRLQSDYQNI
            KSANDYDPTKQNOFWALDARHSVEGIEVYVYKPEFGEKMAEKDRHEQLVK
            EMLYKFSQMKANQSNFIPALDELMSLGKHIDEEOHDIPELEKHLSEESLHMASSF
            ERTKKFVPTHSHPSAPNKPPTETVAGLFAAIDIKLRDMKEKP"
            16383..18905
            /gene="SPAC869.00c"
            16383..18905
            /gene="SPAC869.00c"
            len:116, SIMILARITY:Schizosaccharomyces
            pombe, YAA8_SCHPO, hypothetical 15.4 kd protein c22g7.11c
            in chromosome I., (140 aa), fasta scores: Opt: 192,
            E():2.5e-06, (39.6% identity in 139 aa)"
            /codon_start=1
            /label="SPAC869.00c"
            /product="hypothetical protein"
            /protein_id="CAB60005.1"
            /db_xref="GI:6224586"
            translation="MLOYSKKKVSLNFPFRLSVKMTFRISDAIFKHRLQSDYQNI
            KSANDYDPTKQNOFWALDARHSVEGIEVYVYKPEFGEKMAEKDRHEQLVK
            EMLYKFSQMKANQSNFIPALDELMSLGKHIDEEOHDIPELEKHLSEESLHMASSF
            ERTKKFVPTHSHPSAPNKPPTETVAGLFAAIDIKLRDMKEKP"
            16383..18905
            /gene="SPAC869.00c"
            16383..18905
            /gene="SPAC869.00c"
            len:116, SIMILARITY:Schizosaccharomyces
            pombe, YAA8_SCHPO, hypothetical 15.4 kd protein c22g7.11c
            in chromosome I., (140 aa), fasta scores: Opt: 192,
            E():2.5e-06, (39.6% identity in 139 aa)"
            /codon_start=1
            /label="SPAC869.00c"
            /product="hypothetical protein"
            /protein_id="CAB60004.1"
            /db_xref="GI:6224585"
            translation="MLOYSKKKVSLNFPFRLSVKMTFRISDAIFKHRLQSDYQNI
            KSANDYDPTKQNOFWALDARHSVEGIEVYVYKPEFGEKMAEKDRHEQLVK
            EMLYKFSQMKANQSNFIPALDELMSLGKHIDEEOHDIPELEKHLSEESLHMASSF
            ERTKKFVPTHSHPSAPNKPPTETVAGLFAAIDIKLRDMKEKP"
            16383..18905
            /gene="SPAC869.00c"
            16383..18905
            /gene="SPAC869.00c"
            len:116, SIMILARITY:Schizosaccharomyces
            pombe, YAA8_SCHPO, hypothetical 15.4 kd protein c22g7.11c
            in chromosome I., (140 aa), fasta scores: Opt: 192,
            E():2.5e-06, (39.6% identity in 139 aa)"
            /codon_start=1
            /label="SPAC869.00c"
            /product="hypothetical protein"
            /protein_id="CAB60003.1"
            /db_xref="GI:6224584"
            translation="MLOYSKKKVSLNFPFRLSVKMTFRISDAIFKHRLQSDYQNI
            KSANDYDPTKQNOFWALDARHSVEGIEVYVYKPEFGEKMAEKDRHEQLVK
            EMLYKFSQMKANQSNFIPALDELMSLGKHIDEEOHDIPELEKHLSEESLHMASSF
            ERTKKFVPTHSHPSAPNKPPTETVAGLFAAIDIKLRDMKEKP"
            16383..18905
            /gene="SPAC869.00c"
            16383..18905
            /gene="SPAC869.00c"
            len:116, SIMILARITY:Schizosaccharomyces
            pombe, YAA8_SCHPO, hypothetical 15.4 kd protein c22g7.11c
            in chromosome I., (140 aa), fasta scores: Opt: 192,
            E():2.5e-06, (39.6% identity in 139 aa)"
            /codon_start=1
            /label="SPAC869.00c"
            /product="hypothetical protein"
            /protein_id="CAB60002.1"
            /db_xref="GI:6224583"
            translation="MLOYSKKKVSLNFPFRLSVKMTFRISDAIFKHRLQSDYQNI
            KSANDYDPTKQNOFWALDARHSVEGIEVYVYKPEFGEKMAEKDRHEQLVK
            EMLYKFSQMKANQSNFIPALDELMSLGKHIDEEOHDIPELEKHLSEESLHMASSF
            ERTKKFVPTHSHPSAPNKPPTETVAGLFAAIDIKLRDMKEKP"
            16383..18905
            /gene="SPAC869.00c"
            16383..18905
            /gene="SPAC869.00c"
            len:116, SIMILARITY:Schizosaccharomyces
            pombe, YAA8_SCHPO, hypothetical 15.4 kd protein c22g7.11c
            in chromosome I., (140 aa), fasta scores: Opt: 192,
            E():2.5e-06, (39.6% identity in 139 aa)"
            /codon_start=1
            /label="SPAC869.00c"
            /product="hypothetical protein"
            /protein_id="CAB60001.1"
            /db_xref="GI:6224582"
            translation="MLOYSKKKVSLNFPFRLSVKMTFRISDAIFKHRLQSDYQNI
            KSANDYDPTKQNOFWALDARHSVEGIEVYVYKPEFGEKMAEKDRHEQLVK
            EMLYKFSQMKANQSNFIPALDELMSLGKHIDEEOHDIPELEKHLSEESLHMASSF
            ERTKKFVPTHSHPSAPNKPPTETVAGLFAAIDIKLRDMKEKP"
            16383..18905
            /gene="SPAC869.00c"
            16383..18905
            /gene="SPAC869.00c"
            len:116, SIMILARITY:Schizosaccharomyces
            pombe, YAA8_SCHPO, hypothetical 15.4 kd protein c22g7.11c
            in chromosome I., (140 aa), fasta scores: Opt: 192,
            E():2.5e-06, (39.6% identity in 139 aa)"
            /codon_start=1
            /label="SPAC869.00c"
            /product="hypothetical protein"
            /protein_id="CAB60000.1"
            /db_xref="GI:6224581"
            translation="MLOYSKKKVSLNFPFRLSVKMTFRISDAIFKHRLQSDYQNI
            KSANDYDPTKQNOFWALDARHSVEGIEVYVYKPEFGEKMAEKDRHEQLVK
            EMLYKFSQMKANQSNFIPALDELMSLGKHIDEEOHDIPELEKHLSEESLHMASSF
            ERTKKFVPTHSHPSAPNKPPTETVAGLFAAIDIKLRDMKEKP"
            16383..18905
            /gene="SPAC869.00c"
            16383..18905
            /gene="SPAC869.00c"
            len:116, SIMILARITY:Schizosaccharomyces
            pombe, YAA8_SCHPO, hypothetical 15.4 kd protein c22g7.11c
            in chromosome I., (140 aa), fasta scores: Opt: 192,
            E():2.5e-06, (39.6% identity in 139 aa)"
            /codon_start=1
            /label="SPAC869.00c"
            /product="hypothetical protein"
            /protein_id="CAB60000.1"
            /db_xref="GI:6224581"
            translation="MLOYSKKKVSLNFPFRLSVKMTFRISDAIFKHRLQSDYQNI
            KSANDYDPTKQNOFWALDARHSVEGIEVYVYKPEFGEKMAEKDRHEQLVK
            EMLYKFSQMKANQSNFIPALDELMSLGKHIDEEOHDIPELEKHLSEESLHMASSF
            ERTKKFVPTHSHPSAPNKPPTETVAGLFAAIDIKLRDMKEKP"
            16383..18905
            /gene="SPAC869.00c"
            16383..18905
            /gene="SPAC869.00c"
            len:116, SIMILARITY:Schizosaccharomyces
            pombe, YAA8_SCHPO, hypothetical 15.4 kd protein c22g7.11c
            in chromosome I., (140 aa), fasta scores: Opt: 192,
            E():2.5e-06, (39.6% identity in 139 aa)"
            /codon_start=1
            /label="SPAC869.00c"
            /product="hypothetical protein"
            /protein_id="CAB60000.1"
            /db_xref="GI:6224581"
            translation="MLOYSKKKVSLNFPFRLSVKMTFRISDAIFKHRLQSDYQNI
            KSANDYDPTKQNOFWALDARHSVEGIEVYVYKPEFGEKMAEKDRHEQLVK
            EMLYKFSQMKANQSNFIPALDELMSLGKHIDEEOHDIPELEKHLSEESLHMASSF
            ERTKKFVPTHSHPSAPNKPPTETVAGLFAAIDIKLRDMKEKP"
            16383..18905
            /gene="SPAC869.00c"
            16383..18905
            /gene="SPAC869.00c"
            len:116, SIMILARITY:Schizosaccharomyces
            pombe, YAA8_SCHPO, hypothetical 15.4 kd protein c22g7.11c
            in chromosome I., (140 aa), fasta scores: Opt: 192,
            E():2.5e-06, (39.6% identity in 139 aa)"
            /codon_start=1
            /label="SPAC869.00c"
            /product="hypothetical protein"
            /protein_id="CAB60000.1"
            /db_xref="GI:6224581"
            translation="MLOYSKKKVSLNFPFRLSVKMTFRISDAIFKHRLQSDYQNI
            KSANDYDPTKQNOFWALDARHSVEGIEVYVYKPEFGEKMAEKDRHEQLVK
            EMLYKFSQMKANQSNFIPALDELMSLGKHIDEEOHDIPELEKHLSEESLHMASSF
            ERTKKFVPTHSHPSAPNKPPTETVAGLFAAIDIKLRDMKEKP"
            16383..18905
            /gene="SPAC869.00c"
            16383..18905
            /gene="SPAC869.00c"
            len:116, SIMILARITY:Schizosaccharomyces
            pombe, YAA8_SCHPO, hypothetical 15.4 kd protein c22g7.11c
            in chromosome I., (140 aa), fasta scores: Opt: 192,
            E():2.5e-06, (39.6% identity in 139 aa)"
            /codon_start=1
            /label="SPAC869.00c"
            /product="hypothetical protein"
            /protein_id="CAB60000.1"
            /db_xref="GI:6224581"
            translation="MLOYSKKKVSLNFPFRLSVKMTFRISDAIFKHRLQSDYQNI
            KSANDYDPTKQNOFWALDARHSVEGIEVYVYKPEFGEKMAEKDRHEQLVK
            EMLYKFSQMKANQSNFIPALDELMSLGKHIDEEOHDIPELEKHLSEESLHMASSF
            ERTKKFVPTHSHPSAPNKPPTETVAGLFAAIDIKLRDMKEKP"
            16383..18905
            /gene="SPAC869.00c"
            16383..18905
            /gene="SPAC869.00c"
            len:116, SIMILARITY:Schizosaccharomyces
            pombe, YAA8_SCHPO, hypothetical 15.4 kd protein c22g7.11c
            in chromosome I., (140 aa), fasta scores: Opt: 192,
            E():2.5e-06, (39.6% identity in 139 aa)"
            /codon_start=1
            /label="SPAC869.00c"
            /product="hypothetical protein"
            /protein_id="CAB60000.1"
            /db_xref="GI:6224581"
            translation="MLOYSKKKVSLNFPFRLSVKMTFRISDAIFKHRLQSDYQNI
            KSANDYDPTKQNOFWALDARHSVEGIEVYVYKPEFGEKMAEKDRHEQLVK
            EMLYKFSQMKANQSNFIPALDELMSLGKHIDEEOHDIPELEKHLSEESLHMASSF
            ERTKKFVPTHSHPSAPNKPPTETVAGLFAAIDIKLRDMKEKP"
            16383..18905
            /gene="SPAC869.00c"
            16383..18905
            /gene="SPAC869.00c"
            len:116, SIMILARITY:Schizosaccharomyces
            pombe, YAA8_SCHPO, hypothetical 15.4 kd protein c22g7.11c
            in chromosome I., (140 aa), fasta scores: Opt: 192,
            E():2.5e-06, (39.6% identity in 139 aa)"
            /codon_start=1
            /label="SPAC869.00c"
            /product="hypothetical protein"
            /protein_id="CAB60000.1"
            /db_xref="GI:6224581"
            translation="MLOYSKKKVSLNFPFRLSVKMTFRISDAIFKHRLQSDYQNI
            KSANDYDPTKQNOFWALDARHSVEGIEVYVYKPEFGEKMAEKDRHEQLVK
            EMLYKFSQMKANQSNFIPALDELMSLGKHIDEEOHDIPELEKHLSEESLHMASSF
            ERTKKFVPTHSHPSAPNKPPTETVAGLFAAIDIKLRDMKEKP"
            16383..18905
            /gene="SPAC869.00c"
            16383..18905
            /gene="SPAC869.00c"
            len:116, SIMILARITY:Schizosaccharomyces
            pombe, YAA8_SCHPO, hypothetical 15.4 kd protein c22g7.11c
            in chromosome I., (140 aa), fasta scores: Opt: 192,
            E():2.5e-06, (39.6% identity in 139 aa)"
            /codon_start=1
            /label="SPAC869.00c"
            /product="hypothetical protein"
            /protein_id="CAB60000.1"
            /db_xref="GI:6224581"
            translation="MLOYSKKKVSLNFPFRLSVKMTFRISDAIFKHRLQSDYQNI
            KSANDYDPTKQNOFWALDARHSVEGIEVYVYKPEFGEKMAEKDRHEQLVK
            EMLYKFSQMKANQSNFIPALDELMSLGKHIDEEOHDIPELEKHLSEESLHMASSF
            ERTKKFVPTHSHPSAPNKPPTETVAGLFAAIDIKLRDMKEKP"
            16383..18905
            /gene="SPAC869.00c"
            16383..18905
            /gene="SPAC869.00c"
            len:116, SIMILARITY:Schizosaccharomyces
            pombe, YAA8_SCHPO, hypothetical 15.4 kd protein c22g7.11c
            in chromosome I., (140 aa), fasta scores: Opt: 192,
            E():2.5e-06, (39.6% identity in 139 aa)"
            /codon_start=1
            /label="SPAC869.00c"
            /product="hypothetical protein"
            /protein_id="CAB60000.1"
            /db_xref="GI:6224581"
            translation="MLOYSKKKVSLNFPFRLSVKMTFRISDAIFKHRLQSDYQNI
            KSANDYDPTKQNOFWALDARHSVEGIEVYVYKPEFGEKMAEKDRHEQLVK
            EMLYKFSQMKANQSNFIPALDELMSLGKHIDEEOHDIPELEKHLSEESLHMASSF
            ERTKKFVPTHSHPSAPNKPPTETVAGLFAAIDIKLRDMKEKP"
            16383..18905
            /gene="SPAC869.00c"
            16383..18905
            /gene="SPAC869.00c"
            len:116, SIMILARITY:Schizosaccharomyces
            pombe, YAA8_SCHPO, hypothetical 15.4 kd protein c22g7.11c
            in chromosome I., (140 aa), fasta scores: Opt: 192,
            E():2.5e-06, (39.6% identity in 139 aa)"
            /codon_start=1
            /label="SPAC869.00c"
            /product="hypothetical protein"
            /protein_id="CAB60000.1"
            /db_xref="GI:6224581"
            translation="MLOYSKKKVSLNFPFRLSVKMTFRISDAIFKHRLQSDYQNI
            KSANDYDPTKQNOFWALDARHSVEGIEVYVYKPEFGEKMAEKDRHEQLVK
            EMLYKFSQMKANQSNFIPALDELMSLGKHIDEEOHDIPELEKHLSEESLHMASSF
            ERTKKFVPTHSHPSAPNKPPTETVAGLFAAIDIKLRDMKEKP"
            16383..18905
            /gene="SPAC869.00c"
            16383..18905
            /gene="SPAC869.00c"
            len:116, SIMILARITY:Schizosaccharomyces
            pombe, YAA8_SCHPO, hypothetical 15.4 kd protein c22g7.11c
            in chromosome I., (140 aa), fasta scores: Opt: 192,
            E():2.5e-06, (39.6% identity in 139 aa)"
            /codon_start=1
            /label="SPAC869.00c"
            /product="hypothetical protein"
            /protein_id="CAB60000.1"
            /db_xref="GI:6224581"
            translation="MLOYSKKKVSLNFPFRLSVKMTFRISDAIFKHRLQSDYQNI
            KSANDYDPTKQNOFWALDARHSVEGIEVYVYKPEFGEKMAEKDRHEQLVK
            EMLYKFSQMKANQSNFIPALDELMSLGKHIDEEOHDIPELEKHLSEESLHMASSF
            ERTKKFVPTHSHPSAPNKPPTETVAGLFAAIDIKLRDMKEKP"
            16383..18905
            /gene="SPAC869.00c"
            16383..18905
            /gene="SPAC869.00c"
            len:116, SIMILARITY:Schizosaccharomyces
            pombe, YAA8_SCHPO, hypothetical 15.4 kd protein c22g7.11c
            in chromosome I., (140 aa), fasta scores: Opt: 192,
            E():2.5e-06, (39.6% identity in 139 aa)"
            /codon_start=1
            /label="SPAC869.00c"
            /product="hypothetical protein"
            /protein_id="CAB60000.1"
            /db_xref="GI:6224581"
            translation="MLOYSKKKVSLNFPFRLSVKMTFRISDAIFKHRLQSDYQNI
            KSANDYDPTKQNOFWALDARHSVEGIEVYVYKPEFGEKMAEKDRHEQLVK
            EMLYKFSQMKANQSNFIPALDELMSLGKHIDEEOHDIPELEKHLSEESLHMASSF
            ERTKKFVPTHSHPSAPNKPPTETVAGLFAAIDIKLRDMKEKP"
            16383..18905
            /gene="SPAC869.00c"
            16383..18905
            /gene="SPAC869.00c"
            len:116, SIMILARITY:Schizosaccharomyces
            pombe, YAA8_SCHPO, hypothetical 15.4 kd protein c22g7.11c
            in chromosome I., (140 aa), fasta scores: Opt: 192,
            E():2.5e-06, (39.6% identity in 139 aa)"
```


gene
AEMIQDEVVVKYKIVIRGDNVLAISPTTE"
5646. .5834
/gene="rpl37E"
CDS
5646. .5834
/gene="rpl37E"
/codon_start=1
/transl_table=11
/product="LSU ribosomal protein L37E (rpl37E)"
/protein_id="CAB49564.1"
/db_xref="GI:5458075"
/translation="MSSGPESLGRNKTPTTHRCRCGRKAENVKVKKYCAACGFGGRSR
RLKRYRWSKWKKKKNVH"
complement(5846. .7228)
/note="PAB1937"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB49565.1"
/db_xref="GI:5458076"
/translation="MSLEALFNPKSIATVIGASSKPKIGIVAIMKNLIEYGEKKIYPV
NIRGEMEISGRKFKVYSILEVDEVDMAIIVPAKFPVQVVEECGQKGVKVPPIIS
SGFELGEECKVQEOOLVETARKYGMRIILGNIFGVYTPAKLNGTFGTVDVLPGLA
LISQGLGALMGWTLLEKIGLSAVSVSGNKADIDDAOLLEFFKDDENTRAIIYYME
GVKDRFRMEVAKVSKKPLIIVIKAGRSRGAKAAASHTGSLAGSDTIYSAAPKQSG
VLRALTIGEAFDWARALSNLPEQGDNVLIITNGGGIGVMATDAEEGILKLVLDLEQ
LKIPANHPFGSYKNPVDLTGMADAKSYEGALINDLNHPHMSIIVLYCOTAVLDPR
LDLAEIIRREYNSGRKKPLVVAIVGGIEAKEAIDTUNENIGIPAYPEPERAVKALSALI
RWSKWKAKHK"
complement(7322. .8233)
/note="PAB1936"
/codon_start=1
/transl_table=11
/product="dihydroorotase dehydrogenase (pyrD)"
/protein_id="CAB49566.1"
/db_xref="GI:5458077"
/translation="MVGMLEVNILFGIKFNPLILASGVVDMTPPELLRRHREGAGGV
VTKSIGPRKGYENPTIVELPYGLINAMGLPNPGEAFLEEFKEKRFDPFVIVSIFG
GTPEFAFLSKLEGVADAFELNLSCHPAKGYMEIGOKPENYEVVKAVDKVDKPV
IAKLPVNSDIREGLAKKAGAGVSNITVKAIDIVAKRPIILNKGSGVSGPV
KPIALRAVYDLASSDDIPVIGIGITWQDAVEFLLAGASALQIGTAVYLRGFSVFRE
IAEGISRYLKEEGYSYVKEIITGLAKV"
8283. .9560
/gene="neac"
8283. .9560
/gene="neac"
/note="PAB0443"
/codon_start=1
/transl_table=11
/product="chlorohydrolase, putative"
/protein_id="CAB49567.1"
/db_xref="GI:5458078"
/translation="MVSLSVLLRGGLVLYGRDYKLTADVLIEGDKIVYKRNINKPA
DEVLVSKSLVIPSINATHSPVILRLGLAEDVPLMELWLEYINPVRKLRKDDVYM
GSKLALIEAHSGSTFVDVYFMHEIEAKATEEVEGLRGLVLDLDEKRLKLEMR
ETEKLYEFIKLDSKVNFILAPHAPYTCFSDCLRWSEKSEWNSLVTHLAETQDE
IKIITREKYSKPDVLEVDYGLNLEKILIAHGLIWSDEDIRKISSAGATIAHCPASNNK
LGSQVFPKKALENNVVALGTDGAASNTLDIRMELASILQKVIHRDPSIVKSED
IFPMAILNGAKALGLKAGVIRAGYLADIADLRKAHLPLVNSPLASIIIFSARGGDVD
TLIVGGEIVMLDSLELTVDEEKVIDKLEVSVA"
9557. .10516
/note="PAB0444"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB49568.1"
/db_xref="GI:5458079"
/translation="MIPFKEVLMGIPELVIVVLFYHYHHSRNSSYLMYSPFLSL
GVGIVVEDYLAGLAFSSFFWIGVQLFRRUGGFKGNLKLATVATSLASYNFPL
DRSEVITLVISLMTLLAIMLAFKKVLTISYFLLVLVLSYGLKLVFLVIA
TTLGYELIVETKLTILSPSTGEELKLPKVIFAEITPEDIKLVLFSPKPGK
GERFWVTKLSSDVTNIEPTNLPKTLISLAVDYLAKKGVVVIDCLDYLIENGFESVI
KFWAHLRDYAVHMGSSVITLGPLLEGLSDKEKVMKRALGEEDV"

CDS
10500. .11060
/note="PAB0445"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB49569.1"
/db_xref="GI:5458080"
/translation="MRKMFVHVHGDITRFKAEALVNAANKYLHGGGVAYAIKAASG
DVSEVIRISKEEMKQICGRDIEHGEVVVTPNLAKNGKVIYIHTVGPYGGKWD
KRKLEALILGALKADELGVRSIAFPAISAGIYGCPLLEEVVTKLVVNEFLKSAKN
VTDVTLVLYSERDEYEVALKVLERDEL"
Query Match 2.9%; Score 41.2; DB 1; Length 307120;
Best Local Similarity 53.8%; Pred. No. 1.8;
Matches 85; Conservative 0; Mismatches 73; Indels 0; Gaps 0;
QY 743 agccagacaatcacgcgggaatatggtgtagcgatatagaccagcgagtagtattacc 802
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 282090 ACCCCAGAAAACACGCGGAAACATGGACACAAACCTAATCAGAAAGGACCACCAT 282149
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 803 tatctgccgtacgtgcgcgtgagcgccgcctgtttattggtgatgcccatgctgttcag 862
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 282150 TACTTCCCTGCTCTTCCTTGATGGAGCTTACTCTGGCTATAGCGGATTTCATGCCGTAATG 282209
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 863 ggtgatggtgagatttgcggggaccgcgtagagatttgc 900
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 282210 GGAGATGGCGAAGCTCGCGTCTCGGCTTGTGAGGTTTC 282247
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
RESULT 6
NMA222491/c 349061 bp DNA BCT 30-MAR-2000
LOCUS
DEFINITION
Neisseria meningitidis serogroup A strain Z2491 complete genome;
segment 2/7.
ACCESSION
AL162753 AL157959
VERSION
AL162753.2 GI:7379120
KEYWORDS
Neisseria meningitidis.
Neisseria meningitidis.
SOURCE
Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
Neisseria.
REFERENCE
1 (bases 1 to 349061)
Parkhill,J., Achtman,M., James,K.D., Bentley,S.D., Churcher,C.,
Klee,S.R., Morelli,G., Basham,D., Brown,D., Chillingworth,T.,
Davies,R.M., Davis,P., Devlin,K., Feltwell,T., Hamlin,N.,
Holroyd,S., Jagsels,K., Leather,S., Moule,S., Mungall,K.,
Quail,M.A., Rajandream,M.A., Rutherford,K.M., Simmonds,M.,
Skellton,J., Whitehead,S., Spratt,B.G. and Barrall,B.G.
Complete DNA sequence of a serogroup A strain of Neisseria
meningitidis Z2491
Nature 404 (6777), 502-506 (2000)
JOURNAL
MEDLINE
2222556
REFERENCE
2 (bases 1 to 349061)
Parkhill,J.
Direct Submission
Submitted (30-MAR-2000) Submitted on behalf of the Neisseria
Sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA E-mail: parkhillesanger.ac.uk
Notes:
Details of N. meningitidis sequencing at the Sanger Centre are
available on the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/N_meningitidis/).
FEATURES
source
1..349061
/organism="Neisseria meningitidis"
/strain="Z2491"
/db_xref="taxon:487"
/note="serogroup: A"
complement(24. .206)
/note="ATR repeat; hmms hit to HMM ATR (1 - 183), score:
310.39"
/label=ATR
209. .212
repeat_unit
RBS

```

gene
220..681
/ gene="NMA0368"
CDS
220..681
/ gene="NMA0368"
/ note="NMA0368, probable integral membrane protein, len:
153 aa; contains four probable transmembrane domains"
/codon_start=1
/transl_table=11
/product="putative integral membrane protein"
/protein_id="CAB83669.1"
/db_xref="GI:7379121"
/translation="MQEQNRKSSPIYMLIVSYVALWIASLNSVAFYLGNHGSMGLTV
LILGSIFASLDIRCYAVYANYWLAAILVLLALRKVVPVHAAFGLGALVAFSVKAVYV
DEAGNTSDIVYAGFYLMWAAFAVASIGTTFAGNKERKAASAADGTTNDY"
671..1492
/ gene="hemK"
CDS
671..1492
/ gene="hemK"
/ note="NMA0369, hemK, HemK protein, len: 273 aa; similar
to e.g. HEMK_ECOLI P37186 HEMK protein (277 aa), fasta
scores: E(): 0, 42.3% identity in 279 aa overlap. Contains
PS00092 N-6 Adenine-specific DNA methylases signature"
/codon_start=1
/transl_table=11
/product="HemK protein"
/protein_id="CAB83670.1"
/db_xref="GI:7379122"
/translation="MTDFKWLGLSKLPKNEARMLOVYSEYTRVOLLTRGGEMPDEV
RQRADLAORLNGEPVAYILGAREFYGRRTVNPVSLIPRTEHLEAVLARLPEN
GRVWDLGTGSNAVTVALERPDFAFVRASDISPPALETARKNAADLGRVAFYGSWF
DTMPSEGNKDIIVSNPPYIENGDKHLSQGDLPFEPALTDFDGLSCLRTLAQGAP
DRLAEGGFLLEHGFDDQGAARVGVLAENGFSGVETPLDLAGLDRVTLGKYMHLK"
831..840
/ gene="hemK"
misc_feature
/ note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
1162..1171
/ gene="hemK"
misc_feature
/ note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
1187..1207
/ gene="hemK"
misc_feature
/ note="PS00092 N-6 Adenine-specific DNA methylases
signature"
complement(1279..1288)
/ note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
1565..1568
RBS
1578..1596
/ gene="NMA0370"
CDS
1578..2966
/ gene="NMA0370"
/ note="NMA0370, probable integral membrane protein, len:
462 aa; similar to hypothetical proteins e.g. Y325_HAEIN
P44640 hypothetical protein HI0325 (450 aa), fasta scores:
E(): 0, 49.8% identity in 464 aa overlap"
/codon_start=1
/transl_table=11
/product="putative integral membrane protein"
/protein_id="CAB83671.1"
/db_xref="GI:7379123"
/translation="MNADVAVIVMLVLSLRVHVVLSTIGAFVGVAGVAGMPLQNI
AAGQVSQAGIIPFNKLEGKAKIALSYAMLGAFAMATHISGLPQQLAGAVVRKLN
RGMPDSVRSQGVKWLSTITLVGMMSQNIPIPIAFIPMVPPLLLVNRKLRID
RLIACTVIFGLVTVFMFLPYGFALFLNELLLGNTHSAAPOLDVKNINVMMAIPL
GMLAGLLIAFVHYRKPRLYQSNNDATAGNADANRPQFSVRSLSAAVAIVCFALQI
MYEDSLVGLMGLFAFVGMVGLVINDKANDVFGIKRMWVGFIMIAAQGFPAVWNA
TGHIOPLVSSMAIFGNSKGMALAMLVGLITMGISSESTPIIAIYVPLCVGL
GFSPLIATVIVGTALGADGASPDSTLGTPTMGLNADQGDHDIRDSVIPTPIHYNIP
LLIAGWIAAMVL"
2967..3191
/ gene="slyX"
CDS
2967..3191
/ gene="slyX"
/ note="NMA0371, slyX, SLX protein homolog, len: 74 aa;
similar to SLX_HAEIN P44759 SLX protein homolog (73 aa),
fasta scores: E(): 0.0017, 33.8% identity in 74 aa
overlap, and SLX_ECOLI P30857 SLX protein (72 aa), fasta
scores: E(): 0.24, 32.4% identity in 68 aa overlap"
/codon_start=1
/transl_table=11
/product="SLX protein homolog"
/protein_id="CAB83672.1"
/db_xref="GI:7379124"
/translation="MDAVQEFHRITEIQSALQEDVIAGLNAMVAELRQTLDQQQ
QLRLLYQKMDRNPDAQEPYSLRDEIPHY"
complement(3271..3423)
/ gene="NMA0372"
CDS
complement(3271..3423)
/ gene="NMA0372"
/ note="NMA0372, unknown, questionable CDS, len: 50 aa"
/codon_start=1
/transl_table=11
/product="very hypothetical protein NMA0372"
/protein_id="CAB83673.1"
/db_xref="GI:7379125"
/translation="MFGRLSPLGSDGFFLAVIYEPVGFICLOGKTAIFYRSGGKRFRHI
KGILSD"
3562..3571
/ note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(3586..4356)
/ gene="thiF"
CDS
complement(3586..4356)
/ gene="thiF"
/ note="NMA0373, thiF, probable ThiF protein, len: 256 aa;
similar to e.g. THIF_ECOLI P30138 THIF protein (251 aa),
fasta scores: E(): 0, 43.1% identity in 246 aa overlap,
and MOEB_ECOLI P12282 molybdopterin biosynthesis MOEB
protein. (249 aa), fasta scores: E(): 0, 43.9% identity in
244 aa overlap (note that N.m. does not have orthologs of
any other molybdopterin biosynthesis proteins). Contains
pfam match to entry PF00899 ThiF_family, ThiF family"
/codon_start=1
/transl_table=11
/product="ThiF protein"
/protein_id="CAB83674.1"
/db_xref="GI:7379126"
/translation="MTTETHDNDADFLLVSRHLLDEIGEQQKLSAAHILVCGC
GLGAAALPYLAASGIGTLTIADSDTVLHNLQRFADSGDVGKLTALADRLRHIN
HTVDVITNEKLDGCRLTGLVQRAADIVLCCDNATPQAVNRACVQAKTPLYSGAAVR
FEGQLAVYRPDLDPSPCYACLFDSGSDGICSLGFGVSPGLVIGTGSTQAAEALKILL
DAGEPSHGRVAVYRALEGGWQYFDLPNRPCEPCVGAER"
3787..3796
/ note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(3847..4257)
/ gene="thiF"
misc_feature
/ note="Pfam match to entry PF00899 ThiF_family, ThiF
family, score 186.60, E-value 4.1e-52"
complement(4418..4427)
/ note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
4471..7224
/ gene="ppc"
CDS
4471..7224
/ gene="ppc"
/BC_number="4.1.1.31"
/ note="NMA0374, ppc, phosphoenolpyruvate carboxylase, len:
917 aa; similar to many e.g. CAPP_RHOP A032483
phosphoenolpyruvate carboxylase (EC 4.1.1.31) (936 aa),
fasta scores: E(): 0, 43.3% identity in 928 aa overlap.
Contains 2x Pfam match to entry PF00311 pepcase,
Phosphoenolpyruvate carboxylase, PS00017 ATP/GTP-binding
site motif A (P-loop), and PS00393 Phosphoenolpyruvate
carboxylase active site 2"

```



```
/codon_start=1
/transl_table=11
/product="phosphoenolpyruvate carboxylase"
/protein_id="CAB83675.1"
/db_xref="GI:7379127"
/translation="WRHLRYIYISYGNRKIMOLHILNPNKDAALAAAEFLKOSLFN
LHHEASPLVETVKLLSTSDSAALEKVLPODEQOHDUTLACGLFAQLILNIAED
VHHERRQLHEAGRGAGSLSITVRLKAGKADGKSVQROLDNTSCFVAVLTARHTE
VORQVLNFRIRALLPQRECTNADALRLREIDTILGLMWTSTRRLKLSFVND
ETNNGVSIFPKSFFKLPKMEHDFQATYPOVRVPOILKIGWISGDRDGNPFVS
AETLRFARRIADAVREFYRGELDKLYRELPLSIRRVKVGDMVALSKSPDETARA
EPEYRRATAYIMARAMKARGLGCMCKFGELEPYASAQEFDLDDLKLQLSLDNGS

Query Match      2.9%; Score 41.2; DB 2; Length 349061;
Best Local Similarity 53.0%; Pred. No. 1.8;
Matches 88; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 188 gaataacatataaagggttggaagaatccattatgccaacacgggtgtgtggtccggg 247
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 106727 GCATCTTAACAAAACACGCAAAAGCCATGATGACCGAAGCGAGTTTATCCGCGCG 106668

QY 248 cgtaaaccggtaacgcacatccacgtacggaagaatgcaaaaagagtttcattacaccatt 307
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 106667 AGCGAAGCATTTTGAACACATCGAAGACCAATCGACGAAACGGCTGGGATTTCCGAC 106608

QY 308 gcccttattcacaccggtcgtaccatcgacacgtgacccggtgaccgga 353
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 106607 TGCCGGTTTCCGGAACGTCCTGACCATCGAAGCCGAGACGGCA 106562

RESULT 7
AC020050
LOCUS
DEFINITION
Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 92918)
AUTHORS
Adams M. and Venter, J.C.
TITLE
Direct Submission
JOURNAL
Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT
This sequence was identified as CDM:10211861 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source
1..92918
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT 25719 a 20547 c 20610 g 26042 t
ORIGIN

Query Match      2.7%; Score 39.6; DB 43; Length 92918;
Best Local Similarity 61.8%; Pred. No. 4.2;
Matches 63; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 926 ttgatcaagaactggcagctttcttcgccaagaatgcgaataatattattagtt 985
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 49999 TTGCCCATGATCTGGGGTCCCATTTAGCCAAAAGGTAGCGCGAAATGCTCTCAGC 50058

QY 986 attgcagtcacgtccgctggagatgcgacgcgaattcca 1027
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 50059 ATCGGGGCATCAAGTTCCTGTCAGGATGCGCGGAGGGGAGCA 50100
```

RESULT 8

AC008323

LOCUS

DEFINITION

Drosophila melanogaster chromosome 2 clone BACR03G07 (D850) RPCI-98

03.C.7 map 24D-24D strain Y; cn bw sp, *** SEQUENCING IN PROGRESS

***, 92 unordered pieces.

AC008323

AC008323.2 GI:5957898

HTG; HTGS_PHASE1.

fruit fly.

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

1 (bases 1 to 123296)

AUTHORS

Celisner, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,

Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,

Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,

Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,

Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P.,

Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S.,

Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Shnr, E.,

Svirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and

Rubin, G.M.

Direct Submission

Submitted (02-AUG-1999) Drosophila Genome Center, Lawrence Berkeley

Laboratory, MS 64-121, Berkeley, CA 94720, USA

On Feb 11, 2000 this sequence version replaced gi:5670417.

For further information about this sequence, including its location

and relationship to other sequences, please visit our sequence

archive Web site (<http://www.fruitfly.org/sequence/>) or send emailto bdg@fruitfly.berkeley.edu. All contigs in this submission meet

the following cutoffs: length >= 200 bases.

* NOTE: This is a 'working draft' sequence. It currently

* consists of 92 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence.

* as soon as it is available and the accession number will

* be preserved.

* 1

* 840: contig of 840 bp in length

* 841

* 920: gap of unknown length

* 921

* 2166: contig of 1246 bp in length

* 2167

* 2246: gap of unknown length

* 2247

* 2984: contig of 738 bp in length

* 2985

* 3064: gap of unknown length

* 3065

* 3761: contig of 697 bp in length

* 3762

* 3841: gap of unknown length

* 3842

* 4489: contig of 648 bp in length

* 4490

* 4569: gap of unknown length

* 4570

* 5143: contig of 574 bp in length

* 5144

* 5223: gap of unknown length

* 5224

* 5878: contig of 655 bp in length

* 5879

* 5958: gap of unknown length

* 5959

* 6896: contig of 938 bp in length

* 6897

* 6976: gap of unknown length

* 6977

* 8006: contig of 1030 bp in length

* 8007

* 8086: gap of unknown length

* 8087

* 9208: contig of 1122 bp in length

* 9209

* 9210: gap of unknown length

* 9211

* 9212: gap of unknown length

* 9213

* 9214: gap of unknown length

* 9215

* 9216: gap of unknown length

* 9217

* 9218: gap of unknown length

* 9219

* 9220: gap of unknown length

* 9221

* 9222: gap of unknown length

* 9223

* 9224: gap of unknown length

* 9225

* 9226: gap of unknown length

* 9227

* 9228: gap of unknown length

* 9229

* 9230: gap of unknown length

* 9231

* 9232: gap of unknown length

* 9233

* 9234: gap of unknown length

* 9235

* 9236: gap of unknown length

* 9237

* 9238: gap of unknown length

* 9239

* 9240: gap of unknown length

* 9241

* 9242: gap of unknown length

* 9243

* 9244: gap of unknown length

* 9245

* 9246: gap of unknown length

* 9247

* 9248: gap of unknown length

* 9249

* 9250: gap of unknown length

* 9251

* 9252: gap of unknown length

* 9253

* 9254: gap of unknown length

* 9255

* 9256: gap of unknown length

* 9257

* 9258: gap of unknown length

* 9259

* 9260: gap of unknown length

* 9261

* 9262: gap of unknown length

* 9263

* 9264: gap of unknown length

* 9265

* 9266: gap of unknown length

* 9267

* 9268: gap of unknown length

* 9269

* 9270: gap of unknown length

* 9271

* 9272: gap of unknown length

* 9273

* 9274: gap of unknown length

* 9275

* 9276: gap of unknown length

* 9277

* 9278: gap of unknown length

* 9279

* 9280: gap of unknown length

* 9281

* 9282: gap of unknown length

* 9283

* 9284: gap of unknown length

* 9285

* 9286: gap of unknown length

* 9287

* 9288: gap of unknown length

* 9289

* 9290: gap of unknown length

* 9291

* 9292: gap of unknown length

* 9293

* 9294: gap of unknown length

* 9295

* 9296: gap of unknown length

* 9297

* 9298: gap of unknown length

* 9299

* 9300: gap of unknown length

* 9301

* 9302: gap of unknown length

* 9303

* 9304: gap of unknown length

* 9305

* 9306: gap of unknown length

* 9307

* 9308: gap of unknown length

* 9309

* 9310: gap of unknown length

* 9311

* 9312: gap of unknown length

* 9313

* 9314: gap of unknown length

* 9315

* 9316: gap of unknown length

* 9317

* 9318: gap of unknown length

* 9319

* 9320: gap of unknown length

* 9321

* 9322: gap of unknown length

* 9323

* 9324: gap of unknown length

* 9325

* 9326: gap of unknown length

* 9327

* 9328: gap of unknown length

* 9329

* 9330: gap of unknown length

* 9331

* 9332: gap of unknown length

* 9333

* 9334: gap of unknown length

* 9335

* 9336: gap of unknown length

* 9337

* 9338: gap of unknown length

* 9339

* 9340: gap of unknown length

* 9341

* 9342: gap of unknown length

* 9343

* 9344: gap of unknown length

* 9345

* 9346: gap of unknown length

* 9347

* 9348: gap of unknown length

* 9349

* 9350: gap of unknown length

* 9351

* 9352: gap of unknown length

* 9353

* 9354: gap of unknown length

* 9355

* 9356: gap of unknown length

* 9357

* 9358: gap of unknown length

* 9359

* 9360: gap of unknown length

* 9361

* 9362: gap of unknown length

* 9363

* 9364: gap of unknown length

* 9365

* 9366: gap of unknown length

* 9367

* 9368: gap of unknown length

* 9369

* 9370: gap of unknown length

* 9371

* 9372: gap of unknown length

* 9373

* 9374: gap of unknown length

* 9375

* 9376: gap of unknown length

* 9377

* 9378: gap of unknown length

* 9379

* 9380: gap of unknown length

* 9381

* 9382: gap of unknown length

* 9383

* 9384: gap of unknown length

* 9385

* 9386: gap of unknown length

* 9387

* 9388: gap of unknown length

* 9389

* 9390: gap of unknown length

* 9391

* 9392: gap of unknown length

* 9393

* 9394: gap of unknown length

* 9395

52652	53812:	contig	of 1161	bp in length
53813	53892:	gap	of unknown length	
53893	55383:	contig	of 1491	bp in length
55384	55463:	gap	of unknown length	
55464	57318:	contig	of 1855	bp in length
57319	57398:	gap	of unknown length	
57399	58205:	contig	of 1227	bp in length
58206	58705:	gap	of unknown length	
58706	60069:	contig	of 1364	bp in length
60070	60149:	gap	of unknown length	
60150	61376:	contig	of 1227	bp in length
61377	61456:	gap	of unknown length	
61457	62557:	contig	of 1101	bp in length
62558	62637:	gap	of unknown length	
62638	64089:	contig	of 1452	bp in length
64090	64169:	gap	of unknown length	
64170	66406:	contig	of 2237	bp in length
66407	66486:	gap	of unknown length	
66487	68159:	contig	of 1673	bp in length
68160	68239:	gap	of unknown length	
68240	69609:	contig	of 1460	bp in length
69700	69779:	gap	of unknown length	
69780	71812:	contig	of 2033	bp in length
71813	71892:	gap	of unknown length	
71893	73723:	contig	of 1831	bp in length
73724	73803:	gap	of unknown length	
73804	75480:	contig	of 1677	bp in length
75481	75560:	gap	of unknown length	
75561	77830:	contig	of 2270	bp in length
77831	77910:	gap	of unknown length	
77911	80072:	contig	of 2162	bp in length
80073	80152:	gap	of unknown length	
80153	81871:	contig	of 1719	bp in length
81872	81951:	gap	of unknown length	
81952	83955:	contig	of 2004	bp in length
83956	84035:	gap	of unknown length	
84036	86169:	contig	of 2134	bp in length
86170	86249:	gap	of unknown length	
86250	88430:	contig	of 2181	bp in length
88431	88510:	gap	of unknown length	
88511	91286:	contig	of 2776	bp in length
91287	91368:	gap	of unknown length	
91369	95100:	contig	of 3734	bp in length
95101	95180:	gap	of unknown length	
95181	101800:	contig	of 6620	bp in length
101801	101880:	gap	of unknown length	
101881	107135:	contig	of 5255	bp in length
107136	107215:	gap	of unknown length	
107216	107848:	contig	of 633	bp in length
107849	107928:	gap	of unknown length	
107929	108627:	contig	of 699	bp in length
108628	108707:	gap	of unknown length	
108708	109417:	contig	of 710	bp in length
109418	109497:	gap	of unknown length	
109498	110046:	contig	of 549	bp in length
110047	110126:	gap	of unknown length	
110127	110767:	contig	of 641	bp in length
110768	110847:	gap	of unknown length	

Qy	926	ttgatcaagaactggcgcagcttctcgtgccacgaatggagaatgccgaaaaattatgagt	985
Db	86793	TTGCCATGATCTGGGGTCCCATTTAGCCACAAAGGTGTAGCGGAAATGCTCTCAGC	86852
Qy	986	attgcagtcgactccgctgagatgcgcgcgaattgca	1027
Db	86853	ATCGGGGATCAAGTTCCTCTCAGGATCGCGGAGGGGAGCA	86894

9

AC007420

LOCUS AC007420 130583 bp DNA HTG 20-SEP-1999
 DEFINITION Drosophila melanogaster chromosome 2 clone BACR07M10 (D630) RPCI-98
 07.M.10 map 24A-24D strain Y; cn bw sp, *** SEQUENCING IN PROGRESS
 *** 83 unordered pieces.

ACCESSION

VERSION AC007420.3 GI:5670587

KEYWORDS

SOURCE HTG; HTGS_PHASE1.

ORGANISM

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

1 (bases 1 to 130583)
 Celniker, S.E., Aghayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,
 Butenhoff, C., Champe, M., Chavez, C., Chev, M., Ciesiolka, L.,
 Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,
 Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,
 Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P.,
 Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S.,
 Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E.,
 Svirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and
 Rubin, G.M.

Sequencing of Drosophila melanogaster

TITLE

JOURNAL

REFERENCE

AUTHORS

2 (bases 1 to 130583)
 Celniker, S.E., Aghayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,
 Butenhoff, C., Champe, M., Chavez, C., Chev, M., Ciesiolka, L.,
 Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,
 Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,
 Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P.,
 Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S.,
 Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E.,
 Svirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and
 Rubin, G.M.

Direct Submission

TITLE

JOURNAL

COMMENT

Submitted (28-APR-1999) Drosophila Genome Center, Lawrence Berkeley
 Laboratory, MS 64-121, Berkeley, CA 94720, USA
 On Aug 2, 1999 this sequence version replaced gi:5649319.
 For further information about this sequence, including its location
 and relationship to other sequences, please visit our sequence
 archive Web site (<http://www.fruitfly.org/sequence/>) or send email
 to bdgp@fruitfly.berkeley.edu. All contigs in this submission meet
 the following cutoffs: length >= 200 bases.

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 83 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 782: contig of 782 bp in length
 783 862: gap of unknown length
 863 1444: contig of 582 bp in length
 1445 1524: gap of unknown length
 1525 2743: contig of 1219 bp in length
 2744 2823: gap of unknown length
 2824 3716: contig of 893 bp in length
 3717 3796: gap of unknown length
 3797 4604: contig of 808 bp in length
 4605 4684: gap of unknown length
 4685 5390: contig of 706 bp in length
 5391 5470: gap of unknown length
 5471 6048: contig of 578 bp in length
 6049 6128: gap of unknown length
 6129 7034: contig of 906 bp in length
 7035 7114: gap of unknown length
 7115 7794: contig of 680 bp in length
 7795 7874: gap of unknown length
 7875 8799: contig of 925 bp in length
 8800 8879: gap of unknown length
 8880 9777: contig of 898 bp in length

9778 9857: gap of unknown length
 9858 11049: contig of 1192 bp in length
 11050 11129: gap of unknown length
 11130 12047: contig of 918 bp in length
 12048 12127: gap of unknown length
 12128 12931: contig of 804 bp in length
 12932 13011: gap of unknown length
 13012 14016: contig of 1005 bp in length
 14017 14096: gap of unknown length
 14097 15426: contig of 1330 bp in length
 15427 15506: gap of unknown length
 15507 16623: contig of 1117 bp in length
 16624 16703: gap of unknown length
 16704 17527: contig of 824 bp in length
 17528 17607: gap of unknown length
 17608 18878: contig of 1271 bp in length
 18879 18958: gap of unknown length
 18959 19887: contig of 929 bp in length
 19888 19967: gap of unknown length
 19968 21188: contig of 1221 bp in length
 21189 21268: gap of unknown length
 21269 22540: contig of 1272 bp in length
 22541 22620: gap of unknown length
 22621 23401: contig of 781 bp in length
 23402 23481: gap of unknown length
 23482 24740: contig of 1259 bp in length
 24741 24820: gap of unknown length
 24821 25835: contig of 1015 bp in length
 25836 25915: gap of unknown length
 25916 27422: contig of 1507 bp in length
 27423 27502: gap of unknown length
 27503 28793: contig of 1291 bp in length
 28794 28873: gap of unknown length
 28874 30090: contig of 1217 bp in length
 30091 30170: gap of unknown length
 30171 31566: contig of 1396 bp in length
 31567 31646: gap of unknown length
 31647 33826: contig of 2180 bp in length
 33827 33906: gap of unknown length
 33907 35660: contig of 1754 bp in length
 35661 3740: gap of unknown length
 3741 36675: contig of 935 bp in length
 36676 36755: gap of unknown length
 36756 38411: contig of 1656 bp in length
 38412 38491: gap of unknown length
 38492 40297: contig of 1806 bp in length
 40298 40377: gap of unknown length
 40378 41875: contig of 1498 bp in length
 41876 41955: gap of unknown length
 41956 43789: contig of 1834 bp in length
 43790 43869: gap of unknown length
 43870 45471: contig of 1602 bp in length
 45472 45551: gap of unknown length
 45552 47128: contig of 1577 bp in length
 47129 47209: gap of unknown length
 47209 48263: contig of 1055 bp in length
 48264 48343: gap of unknown length
 48344 50567: contig of 2224 bp in length
 50568 50647: gap of unknown length
 50648 53133: contig of 2485 bp in length
 53133 53212: gap of unknown length
 53213 54731: contig of 1519 bp in length
 54732 54811: gap of unknown length
 54812 56680: contig of 1869 bp in length
 56681 56760: gap of unknown length
 56761 58314: contig of 1554 bp in length
 58315 58394: gap of unknown length
 58395 60268: contig of 1874 bp in length
 60269 60348: gap of unknown length
 60349 62780: contig of 2432 bp in length
 62781 62860: gap of unknown length
 62861 64898: contig of 2038 bp in length
 64899 64978: gap of unknown length

* 64979	66963:	contig of 1985 bp in length
* 66964	67043:	gap of unknown length
* 67044	69574:	contig of 2531 bp in length
* 69575	69654:	gap of unknown length
* 69655	72365:	contig of 2711 bp in length
* 72366	72445:	gap of unknown length
* 72446	75169:	contig of 2724 bp in length
* 75170	75249:	gap of unknown length
* 75250	78381:	contig of 3132 bp in length
* 78382	78461:	gap of unknown length
* 81086	81086:	contig of 2625 bp in length
* 81166	81166:	gap of unknown length
* 83892	83892:	contig of 2726 bp in length
* 83972	83972:	gap of unknown length
* 86588	86588:	contig of 2616 bp in length
* 86589	86588:	gap of unknown length
* 86669	91696:	contig of 5028 bp in length
* 91697	91776:	gap of unknown length
* 91777	94463:	contig of 2687 bp in length
* 94464	94543:	gap of unknown length
* 94544	98576:	contig of 4033 bp in length
* 98577	98656:	gap of unknown length
* 98657	101629:	contig of 2973 bp in length
* 101630	101709:	gap of unknown length
* 101710	112556:	contig of 10767 bp in length
* 112557	112557:	contig of 901 bp in length
* 113458	113537:	gap of unknown length
* 113538	114183:	contig of 646 bp in length
* 114184	114263:	gap of unknown length
* 114264	114990:	contig of 727 bp in length
* 114991	115070:	gap of unknown length
* 115071	115790:	contig of 720 bp in length
* 115791	115870:	gap of unknown length
* 115871	116529:	contig of 659 bp in length
* 116530	116609:	gap of unknown length
* 116610	116697:	contig of 888 bp in length
* 117498	117577:	gap of unknown length
* 117578	118139:	contig of 562 bp in length
* 118140	118219:	gap of unknown length
* 118220	118669:	contig of 650 bp in length
* 118870	118949:	gap of unknown length
* 118950	119638:	contig of 689 bp in length
* 119639	119718:	gap of unknown length
* 119719	120490:	contig of 772 bp in length
* 120491	120570:	gap of unknown length
* 120571	121288:	contig of 718 bp in length
* 121289	121368:	gap of unknown length
* 121369	122040:	contig of 672 bp in length
* 122041	122121:	gap of unknown length
* 122121	122701:	contig of 581 bp in length
* 122702	122781:	gap of unknown length
* 122782	123465:	contig of 684 bp in length
* 123466	123465:	gap of unknown length
* 123546	124188:	contig of 643 bp in length

Query Match	2.7%;	Score 39.6;	DB 41;	Length 130583;
Best Local Similarity	61.8%;	Pred. No. 4.4;		
Matches 63;	Conservative 0;	Mismatches 39;	Indels 0;	Gaps 0;
Qy 926	ttgatcaagaactgcagcttcttcgtggccacgaatgagaatgcccgaataattattagat	985		
Db 93378	TTGGCCATGATCGGGGTCCCATTAGCCACAAAAGGTAGGCCGCAAAATGCTCTCAGC	93437		
Qy 986	attgagcagtgcacgtccgctgagagatgcagcgaattgca	1027		
Db 93438	ATCGGGGCATCAAGTTCCTGTCGAGGATGCGGAGGGGAGCA	93479		
RESULT 10				
AE003579/c	304383 bp	DNA	INV	22-MAR-2000
LOCUS	AE003579			
DEFINITION	Drosophila melanogaster genomic scaffold 142000013386046	section 5		

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

of 16, complete sequence.
AE003579 AE002638
GI:7295765

HTG.

fruit fly.

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 304383)

Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galie,R.F., George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N., Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.Y., Brandon,R.C., Rogers,Y.H., Blazek,R.G., Champe,M., Pfeiffer,B.D., Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor Miklos,G.L., Abril,J.P., Aghayani,A., An,H.J., Basu,A., Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A., Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y., Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkov,D., Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C., Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I., Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M., Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C., Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferriera,S., Fleischmann,W., Fosler,C., Gabriellian,A.E., Garg,N.S., Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z., Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J., Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J., Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z., Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C., Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A., Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C., McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C., Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L., Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K., Nussekern,D.R., Pacle,J.M., Palazzolo,M., Pittman,G.S., Pan,S., Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K., Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I., Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C., Stapleton,M., Strong,R., Sun,E., Swirskas,R., Tector,C., Turner,R., Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A., Weinstein,G.M., Weissbach,J., Williams,S.M., Woodage,T., Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F., Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H., Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O., Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.

The genome sequence of Drosophila melanogaster

Science 287 (5461), 2185-2195 (2000)

2196006

2 (bases 1 to 304383)

Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.

Direct Submission

Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA

Location/Qualifiers

1. 304383

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/chromosome="2L"

complement(<9367..>9540)

/gene="CG15419"

/product="CT35478"

/db_xref="FLYBASE:FBan0015419"

/db_xref="FLYBASE:FBgn0040710"

/evidence=not_experimental

complement(<9367..>9540)

/gene="CG15419"

/db_xref="FLYBASE:FBan0015419"

/db_xref="FLYBASE:FBgn0040710"

/evidence=not_experimental

complement(9367..9540)

CDS

FEATURES

source

mrna

gene

CDS

[illegible][illegible]

KEYWORDS HTG: D22S1152.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 118593)
AUTHORS Burton,J.
TITLE Direct Submission
JOURNAL Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
requests: clonerquest@sanger.ac.uk
COMMENT On Dec 13, 1999 this sequence version replaced gi:1903194.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em.; EMBL; Sw.; SWISSPROT; Tr.; TrEMBL; Wp.; WORMPEP; Information
on the WORMPEP database can be found at
<http://www.sanger.ac.uk/projects/Celegans/wormpep> This sequence
was generated from part of bacterial clone contigs of human
Chromosome 22, constructed by the Sanger Centre Chromosome 22
Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr22>
RP1-288L1 is from the library RPC1-1 constructed at the Roswell
Park Cancer Institute by the group of Pieter de Jong. For further
details see <http://bacpac.med.buffalo.edu/>
VECTOR: pCYPAC2
This sequence is the entire insert of clone RP1-288L1.
FEATURES
Location/Qualifiers
1..118593
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="22"
/clone="RP1-288L1"
/clone_lib="RPC1-1"
4..141
/note="WSTA repeat: matches 185..314 of consensus"
repeat_region
142..852
/note="LIPAB repeat: matches 534..6163 of consensus"
repeat_region
853..1031
/note="WSTA repeat: matches 1..200 of consensus"
repeat_region
1037..1809
/note="HERVL repeat: matches 3871..4641 of consensus"
repeat_region
2248..2413
/note="HERVL repeat: matches 3359..3525 of consensus"
repeat_region
3173..3373
/note="LTR33 repeat: matches 8..223 of consensus"
repeat_region
3673..4275
/note="L1M4 repeat: matches 2681..3306 of consensus"
repeat_region
4325..4939
/note="HERV16 repeat: matches 1665..2289 of consensus"
repeat_region
4975..5684
/note="HERV16 repeat: matches 2344..3072 of consensus"
repeat_region
5924..6339
/note="HERVL repeat: matches 3442..3860 of consensus"
misc_feature
6296..6578
/note="match: STS: Em:F08316"
repeat_region
6452..6536
/note="HERVL repeat: matches 4031..4115 of consensus"
repeat_region
6537..6824
/note="AluSc repeat: matches 13..299 of consensus"
repeat_region
6825..7958
/note="HERVL repeat: matches 4115..5247 of consensus"

repeat_region
8198..8219
/note="11 copies 2 mer aa 100 conserved"
repeat_region
8803..8918
/note="L2 repeat: matches 2624..2748 of consensus"
repeat_region
9068..9178
/note="MER94 repeat: matches 1..113 of consensus"
repeat_region
9957..15197
/note="LIPAB repeat: matches 900..6146 of consensus"
repeat_region
15193..15979
/note="LIPAB repeat: matches 1..776 of consensus"
repeat_region
16282..17315
/note="SVA repeat: matches 14..955 of consensus"
misc_feature
16843..17747
/note="Cpg island"
/evidence="not_experimental"
repeat_region
17316..17875
/note="SVA repeat: matches 798..1363 of consensus"
misc_feature
complement(17752..18176)
/note="match: GSS: Em:AQ113498"
misc_feature
17896..18316
/note="match: GSS: Em:AQ802108"
repeat_region
18158..18396
/note="MIR repeat: matches 15..262 of consensus"
repeat_region
18529..18648
/note="L2 repeat: matches 2630..2749 of consensus"
repeat_region
18682..18805
/note="MIR repeat: matches 52..173 of consensus"
repeat_region
19153..19462
/note="AluX repeat: matches 1..310 of consensus"
repeat_region
20064..20128
/note="LIPAB repeat: matches 6075..6139 of consensus"
repeat_region
20456..20899
/note="L2 repeat: matches 2218..2742 of consensus"
repeat_region
21834..23162
/note="L1MC2 repeat: matches 4679..6064 of consensus"
repeat_region
23163..23463
/note="AluSq repeat: matches 1..302 of consensus"
repeat_region
23464..23735
/note="L1MC2 repeat: matches 6064..6328 of consensus"
repeat_region
23738..23779
/note="21 copies 2 mer aa 76 conserved"
repeat_region
24053..24998
/note="LIPAB repeat: matches 5344..6264 of consensus"
repeat_region
25021..25378
/note="L2 repeat: matches 1660..2020 of consensus"
repeat_region
25440..25665
/note="MIR repeat: matches 8..243 of consensus"
repeat_region
26462..26542
/note="MIR repeat: matches 52..147 of consensus"
repeat_region
26815..27048
/note="MIR repeat: matches 13..251 of consensus"
repeat_region
27272..27533
/note="AluSq repeat: matches 1..264 of consensus"
repeat_region
28220..28347
/note="MIR repeat: matches 18..152 of consensus"
repeat_region
28620..28790
/note="MIR repeat: matches 89..261 of consensus"
repeat_region
28835..28983
/note="MIR repeat: matches 90..231 of consensus"
repeat_region
29619..29735
/note="L2 repeat: matches 2418..2529 of consensus"
repeat_region
31018..31154
/note="L2 repeat: matches 2240..2400 of consensus"
repeat_region
31217..31587
/note="THELC repeat: matches 1..371 of consensus"
repeat_region
31723..31850
/note="L2 repeat: matches 2577..2707 of consensus"
repeat_region
32234..32433
/note="MIR repeat: matches 47..256 of consensus"
repeat_region
32507..32734
/note="MIR repeat: matches 28..262 of consensus"
repeat_region
33524..33626
/note="MIR repeat: matches 23..121 of consensus"

AE002546.1 GI:7227234
 Neisseria meningitidis.
 Neisseria meningitidis
 Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
 Neisseria.
 1 (bases 1 to 10619)
 Tettelin, H., Saunders, N.J., Heidelberg, J., Jeffries, A.C.,
 Nelson, K.E., Eisen, J.A., Ketchum, K.A., Hood, D.W., Peden, J.F.,
 Dodson, R.J., Nelson, W.C., Gwinn, M.L., Deboy, R., Peterson, J.D.,
 Hickey, E.K., Haft, D.H., Salzberg, S.L., White, O., Fleischmann, R.D.,
 Dougherty, B.A., Mason, T., Ciecko, A., Parksey, D.S., Blair, E.,
 Cittiene, H., Clark, E.B., Cotton, M.D., Utterback, T.R., Khouri, H.,
 Qin, H., Vamathevan, J., Gill, J., Scarlato, V., Masigiani, V.,
 Pizza, M., Grandi, G., Sun, L., Smith, H.O., Fraser, C.M., Moxon, E.R.,
 Rappuoli, R. and Venter, J.C.
 Complete genome sequence of *Neisseria meningitidis* serogroup B
 strain MC58
 Science 287 (5459), 1809-1815 (2000)
 20175755
 2 (bases 1 to 10619)
 Tettelin, H., Saunders, N.J., Heidelberg, J., Jeffries, A.C.,
 Nelson, K.E., Eisen, J.A., Ketchum, K.A., Hood, D.W., Peden, J.F.,
 Dodson, R.J., Nelson, W.C., Gwinn, M.L., Deboy, R., Peterson, J.D.,
 Hickey, E.K., Haft, D.H., Salzberg, S.L., White, O., Fleischmann, R.D.,
 Dougherty, B.A., Mason, T., Ciecko, A., Parksey, D.S., Blair, E.,
 Cittiene, H., Clark, E.B., Cotton, M.D., Utterback, T.R., Khouri, H.,
 Qin, H., Vamathevan, J., Gill, J., Scarlato, V., Masigiani, V.,
 Pizza, M., Grandi, G., Sun, L., Smith, H.O., Fraser, C.M., Moxon, E.R.,
 Rappuoli, R. and Venter, J.C.
 Direct Submission
 Submitted (17-MAR-2000) The Institute for Genomic Research, 9712
 Medical Center Dr, Rockville, MD 20850, USA
 Location/Qualifiers
 1..10619
 /organism="Neisseria meningitidis"
 /strain="MC58"
 /db_xref="taxon:487"
 /note="serogroup: B"
 complement(77..367)
 /gene="NMB1973"
 complement(77..367)
 /gene="NMB1973"
 /note="similar to SP:P77913 PID:1513097 percent identity:
 98.96; identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="chaperonin, 10 kDa"
 /protein_id="AAF42302.1"
 /db_xref="GI:7227235"
 /translation="MTIRPLDVRVVRKLEAEKTSAGIVLPGAAREKPDGMGEVIAV
 AGKTKDGRRLPDVKVDGIKIFGYSQTVKADGEELLVMREEDI FGIVEK"
 595..1247
 /gene="NMB1974"
 /note="This region contains a gene with one or more
 premature stops or frame shifts, and is not the result of a
 sequencing artifact; similar to GB:X59756 percent
 identity: 82.61; identified by sequence similarity;
 putative; IS1016C2 transposase, degenerate"
 1460..2995
 /gene="NMB1975"
 1460..2995
 /gene="NMB1975"
 /note="similar to GB:142023 SP:P44849 PID:1005680
 PID:120824 PID:1204983 percent identity: 77.47;
 identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="sodium- and chloride-dependent transporter"
 /protein_id="AAF42303.1"
 /db_xref="GI:7227236"
 /translation="MSDSKLTAEATGTRAFPMIAAIGSAVGLNWRFPYTAFGNG
 GAFILPYVALLTAGIPLLRGSAPLAFRLGRWFEPVGVNMVYTNIL
 GAFILPYVALLTAGIPLLRGSAPLAFRLGRWFEPVGVNMVYTNIL

15566 15585: gap of unknown length
15586 15589: contig of 984 bp in length
15587 15590: gap of unknown length
15590 17562: contig of 973 bp in length
17563 17582: gap of unknown length
17583 18414: contig of 832 bp in length
18415 18434: gap of unknown length
18435 19614: contig of 1180 bp in length
19615 19634: gap of unknown length
19635 20885: contig of 1251 bp in length
20886 20905: gap of unknown length
20906 22136: contig of 1231 bp in length
22137 22156: gap of unknown length
22157 23503: contig of 1347 bp in length
23504 23524: gap of unknown length
23525 24606: contig of 1083 bp in length
24607 24626: gap of unknown length
24627 25765: contig of 1139 bp in length
25766 25785: gap of unknown length
25786 26995: contig of 1110 bp in length
26996 26915: gap of unknown length
26916 28149: contig of 1234 bp in length
28150 28169: gap of unknown length
28170 29503: contig of 1334 bp in length
29504 29523: gap of unknown length
29524 31129: contig of 1606 bp in length
31130 31149: gap of unknown length
31150 32506: contig of 1357 bp in length
32507 32526: gap of unknown length
32527 33151: contig of 625 bp in length
33152 33171: gap of unknown length
33172 34833: contig of 1662 bp in length
34834 34853: gap of unknown length
34854 36463: contig of 1610 bp in length
36464 36483: gap of unknown length
36484 37769: contig of 1286 bp in length
37770 37789: gap of unknown length
37790 39247: contig of 1458 bp in length
39248 39267: gap of unknown length
39268 41092: contig of 1825 bp in length
41093 41112: gap of unknown length
41113 42002: contig of 890 bp in length
42003 42022: gap of unknown length
42024 43449: contig of 1627 bp in length
43450 43669: gap of unknown length
43670 45468: contig of 1799 bp in length
45469 45488: gap of unknown length
45489 46976: contig of 1488 bp in length
46977 46996: gap of unknown length
46997 48341: contig of 1345 bp in length
48342 48361: gap of unknown length
48362 50274: contig of 1913 bp in length
50275 50294: gap of unknown length
50295 52835: contig of 2541 bp in length
52836 52855: gap of unknown length
52856 54462: contig of 1607 bp in length
54463 54482: gap of unknown length
54483 55386: contig of 904 bp in length
55387 55406: gap of unknown length
55407 57489: contig of 2083 bp in length
57490 57509: gap of unknown length
57510 60192: contig of 2683 bp in length
60193 60212: gap of unknown length
60213 62373: contig of 2161 bp in length
62374 64223: gap of unknown length
64224 64243: gap of unknown length
64244 66061: contig of 1818 bp in length
66062 68081: gap of unknown length
68082 68233: contig of 2152 bp in length
68234 68253: gap of unknown length
68254 70076: contig of 1823 bp in length
70077 70096: gap of unknown length

70097 72046: contig of 1950 bp in length
72047 72066: gap of unknown length
72067 74963: contig of 2897 bp in length
74964 74983: gap of unknown length
74984 77410: contig of 2427 bp in length
77411 77430: gap of unknown length
77431 80001: contig of 2571 bp in length
80002 80021: gap of unknown length
80022 81918: contig of 1897 bp in length
81919 81938: gap of unknown length
81939 83880: contig of 1942 bp in length
83881 83900: gap of unknown length
83901 87305: contig of 3405 bp in length
87306 87325: gap of unknown length
87326 90628: contig of 3303 bp in length
90629 90648: gap of unknown length
90649 93277: contig of 2629 bp in length
93278 93297: gap of unknown length
93298 95346: contig of 2029 bp in length
95347 95366: gap of unknown length
95367 98750: contig of 3404 bp in length
98751 98770: gap of unknown length
98771 101763: contig of 2993 bp in length
101764 101783: gap of unknown length
101784 105858: contig of 4075 bp in length
105859 105878: gap of unknown length
105879 108448: contig of 2570 bp in length
108449 108468: gap of unknown length
108469 112321: contig of 3853 bp in length
112322 112341: gap of unknown length
112342 116841: contig of 4500 bp in length

Query Match 2.7%; Score 39.2; DB 51; Length 133190;
Best Local Similarity 49.1%; Pred. No. 5.8;
Matches 104; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

QY 320 acaccgcctgaccatcgaaacccggtgaccggtatttcgacactgacatgcagatgctttt 379
Db 111107 ACCACCGGGTGTGCTATCGTATGGCAGACTGGATTATTTCGTGCTCGAGTGCTATT 111048

QY 380 gaagggtcatcaatcggaaacagatattccgagccagtgctataaaagcccttctc 439
Db 111047 GGAGCTCTTCTTCTTCGGGGCGGATGCGTCCATTGCACTCCGACAGCGCTCCCGGCC 110988

QY 440 aaccacaacacgaccgatcatggtcaatggcgcggagagaagtgatgctgcgtgctc 499
Db 110987 GGCGGTAAAGTTGCTGTAGCGATCGCAAACTCGTCCCAAGTGGCTTCATGATGCTCTG 110928

QY 500 tatatcgatccatgttgcgccgcgcgttga 531
Db 110927 GTACTCGCGTCTATCTCTCGCCGTCGAGGTCA 110896

RESULT 15
AC009369/c

LOCUS AC009369 154381 bp DNA 18-FEB-2000
DEFINITION Drosophila melanogaster clone RPCI98-8G12, *** SEQUENCING IN PROGRESS ***, 74 unordered pieces.
ACCESSION AC009369
VERSION AC009369.5 GI:6996666
KEYWORDS HTG; HTGS_PHASE1.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster

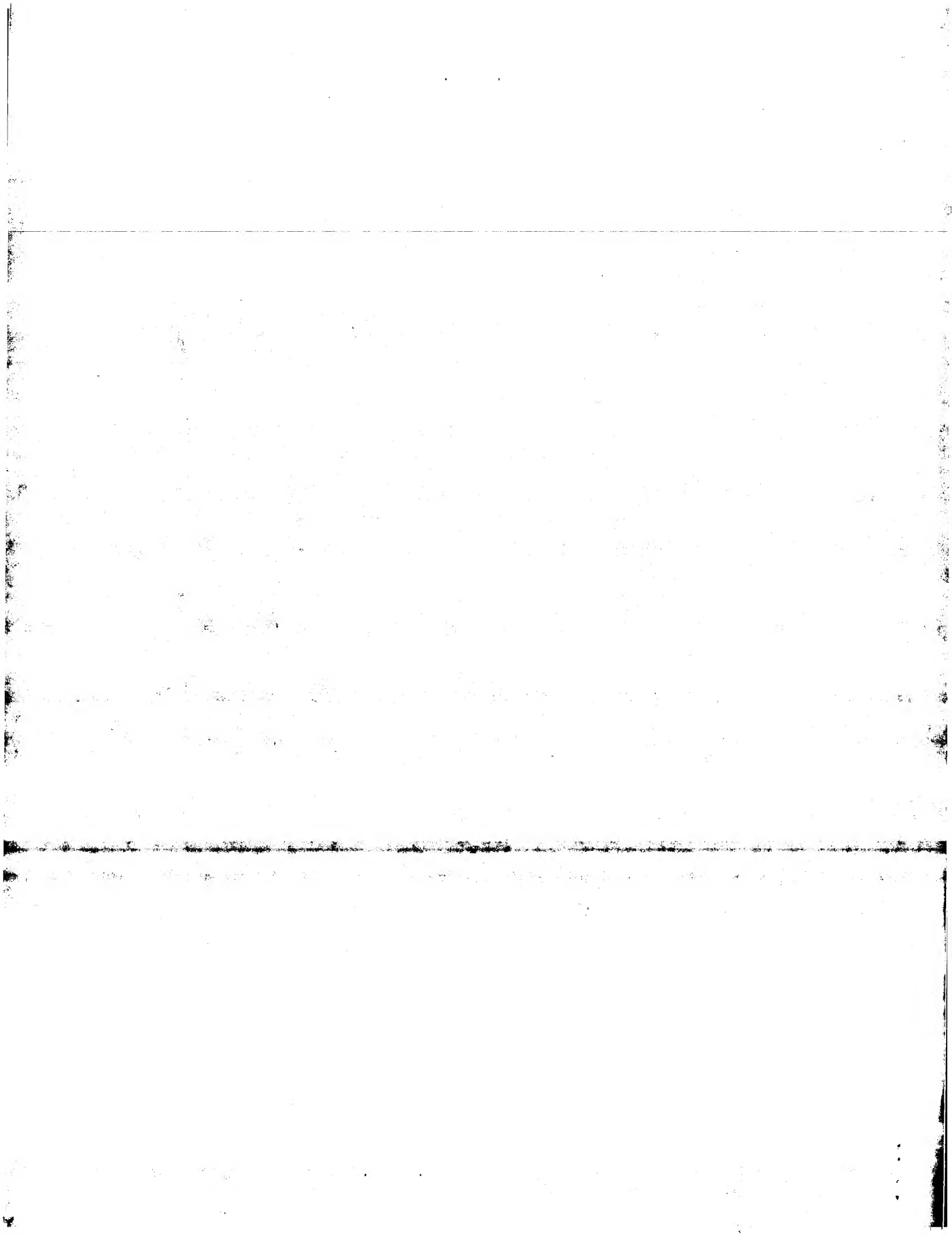
REFERENCE Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 154381)
Muzny,D.M., Adams,C., Bailey,M., Barberia,J., Blankenburg,K., Bodota,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C., Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C., David,R., Delgado,O., Deshazo,D., Ding,Y., Doman-Rashid,N., Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D., Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,

Guevara, W., Harris, K., Hernandez, J., Hodgson, A., Hoques, M.,
 Holloway, C., Hosack, H., Jackson, L.E., Jackson, L., Jia, Y., Jones, M.,
 Kelly, S., Kondejewski, N., Kong, Y., Kovar, C., Leal, B., Li, Z.,
 Lichtarge, O., Liu, J., Liu, W., Logan, O., Lozano, R.J., Lu, J.,
 Lucier, R., Martin, R., Martinez, C., McLeod, M.P., Mei, G., Morgan, M.,
 Morris, S., Nash, S., Nelson, A., Nguyen, R., Nguyen, N., Nguyen, S.,
 Oswal, G., Parish, B., Paxton, S., Payton, B., Perez, L., Pu, L.L.,
 Quiles, M., Reiter, D., Rives, M., Samuel, S., Say, J., Scherer, S.,
 Shah, E., Shen, H., Simon, M., Sparks, A., Stamps, A., Sucgang, R.,
 Tabor, P., Taylor, T., Vasquez, L., Vinson, R., Vo, O., Wahbah, M.,
 Watlington, S., Weinstein, G., Weinstein, I.R., Williamson, A.,
 Worley, K., Wren, J., Wrensford, G., Yu, W., Zhou, X., Nelson, D. and
 Gibbs, R.
 Direct Submission
 Unpublished
 2 (bases 1 to 154381)
 Worley, K.C.
 Direct Submission
 Submitted (20-AUG-1999) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Feb 18, 2000 this sequence version replaced gi:5881432.
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: DRAD
 Center clone name: RPC198-8G12
 ----- Summary Statistics
 Sequencing vector: M13; L08821
 Chemistry: Dye-terminator Big Dye: 4% of reads
 Assembly program: Phrap; version 0.980611
 Consensus quality: 100833 bases at least Q40
 Consensus quality: 122674 bases at least Q30
 Consensus quality: 132575 bases at least Q20
 Estimated insert size: 144035; sum-of-contigs estimation
 Quality coverage: 2.1x in Q20 bases; sum-of-contigs estimation

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 74 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 * 1 782: contig of 782 bp in length
 * 783 802: gap of unknown length
 * 803 1638: contig of 836 bp in length
 * 1639 1658: gap of unknown length
 * 1659 2546: contig of 888-bp in length
 * 2547 2566: gap of unknown length
 * 2567 3735: contig of 1169 bp in length
 * 3736 3755: gap of unknown length
 * 3756 4642: contig of 887 bp in length
 * 4643 4662: gap of unknown length
 * 4663 5583: contig of 921 bp in length
 * 5584 5603: gap of unknown length
 * 5604 6853: contig of 1250 bp in length
 * 6854 6873: gap of unknown length
 * 6874 7895: contig of 1022 bp in length
 * 7896 7915: gap of unknown length
 * 7916 8937: contig of 1022 bp in length
 * 8938 8957: gap of unknown length
 * 8958 10539: contig of 1582 bp in length
 * 10540 10559: gap of unknown length
 * 10560 11902: contig of 1343 bp in length
 * 11903 11922: gap of unknown length
 * 11923 13444: contig of 1522 bp in length
 * 13445 13464: gap of unknown length
 *
 * 13465 14836: contig of 1372 bp in length
 * 14837 14856: gap of unknown length
 * 14857 16113: contig of 1257 bp in length
 * 16114 16133: gap of unknown length
 * 16134 16953: contig of 820 bp in length
 * 16954 16973: gap of unknown length
 * 16974 17949: contig of 976 bp in length
 * 17950 17969: gap of unknown length
 * 17970 19251: contig of 1282 bp in length
 * 19252 19271: gap of unknown length
 * 19272 20147: contig of 876 bp in length
 * 20148 20167: gap of unknown length
 * 20168 21490: contig of 1323 bp in length
 * 21491 21510: gap of unknown length
 * 21511 22371: contig of 861 bp in length
 * 22372 22391: gap of unknown length
 * 22392 23573: contig of 1182 bp in length
 * 23574 23593: gap of unknown length
 * 23594 24778: contig of 1185 bp in length
 * 24779 24798: gap of unknown length
 * 24799 26680: contig of 1882 bp in length
 * 26681 26700: gap of unknown length
 * 26701 27675: contig of 975 bp in length
 * 27676 27695: gap of unknown length
 * 27696 28733: contig of 1038 bp in length
 * 28734 28753: gap of unknown length
 * 28754 30178: contig of 1425 bp in length
 * 30179 30198: gap of unknown length
 * 30199 31626: contig of 1428 bp in length
 * 31627 31646: gap of unknown length
 * 31647 32836: contig of 1190 bp in length
 * 32837 32856: gap of unknown length
 * 32857 34558: contig of 1702 bp in length
 * 34559 34578: gap of unknown length
 * 34579 35924: contig of 1346 bp in length
 * 35925 35944: gap of unknown length
 * 35945 37736: contig of 1792 bp in length
 * 37737 37757: gap of unknown length
 * 37757 38769: contig of 1013 bp in length
 * 38770 38789: gap of unknown length
 * 38790 39846: contig of 1057 bp in length
 * 39847 39866: gap of unknown length
 * 39867 41513: contig of 1647 bp in length
 * 41514 41533: gap of unknown length
 * 41534 43110: contig of 1577 bp in length
 * 43111 43130: gap of unknown length
 * 43131 45511: contig of 2381 bp in length
 * 45512 45531: gap of unknown length
 * 45532 46655: contig of 1124 bp in length
 * 46656 46675: gap of unknown length
 * 46676 47948: contig of 1273 bp in length
 * 47949 47968: gap of unknown length
 * 47969 50813: contig of 2845 bp in length
 * 50814 50833: gap of unknown length
 * 50834 53480: contig of 2647 bp in length
 * 53481 53501: gap of unknown length
 * 53501 54934: contig of 1434 bp in length
 * 54935 54954: gap of unknown length
 * 54955 57412: contig of 2458 bp in length
 * 57413 57432: gap of unknown length
 * 57433 59817: contig of 2385 bp in length
 * 59818 59837: gap of unknown length
 * 59838 61341: contig of 1504 bp in length
 * 61342 61361: gap of unknown length
 * 61362 63044: contig of 1683 bp in length
 * 63045 63064: gap of unknown length
 * 63065 64833: contig of 1769 bp in length
 * 64834 64853: gap of unknown length
 * 64854 67374: contig of 2521 bp in length
 * 67375 67394: gap of unknown length
 * 67395 68813: contig of 1419 bp in length
 * 68814 68833: gap of unknown length
 * 70648: contig of 1815 bp in length

* 70649	70668: gap of unknown length
* 70669	71950: contig of 1282 bp in length
* 71951	71970: gap of unknown length
* 71971	74720: contig of 2750 bp in length
* 74721	74740: gap of unknown length
* 74741	76716: contig of 1976 bp in length
* 76717	76736: gap of unknown length
* 76737	78617: contig of 1881 bp in length
* 78618	78637: gap of unknown length
* 78638	81017: contig of 2380 bp in length
* 81018	81037: gap of unknown length
* 81038	83770: contig of 2733 bp in length
* 83771	83790: gap of unknown length
* 83791	86125: contig of 2335 bp in length
* 86126	86145: gap of unknown length
* 86146	89229: contig of 3084 bp in length
* 89230	89249: gap of unknown length
* 89250	90667: contig of 1418 bp in length
* 90668	90687: gap of unknown length
* 90688	92693: contig of 2006 bp in length
* 92694	92713: gap of unknown length
* 92714	95077: contig of 2364 bp in length
* 95078	95097: gap of unknown length
* 95098	97907: contig of 2810 bp in length
* 97908	97927: gap of unknown length
* 97928	100693: contig of 2766 bp in length
* 100694	100713: gap of unknown length
* 100714	103619: contig of 2906 bp in length
* 103620	103639: gap of unknown length
* 103640	106885: contig of 3246 bp in length
* 106886	106905: gap of unknown length
* 106906	109193: contig of 2288 bp in length
* 109194	109213: gap of unknown length
* 109214	112158: contig of 2945 bp in length
* 112159	112178: gap of unknown length
* 112179	116985: contig of 4807 bp in length
Query Match	
Best Local Similarity 49.1%; Score 39.2; DB 51; Length 154381;	
-Matches 104; Conservative 0; Mismatches 108; Indels 0; Gaps 0;	
QY 320	acaccgctctgaccatcgaccggtgaccggtatttgcacactcgagatgctttt 379
Db 140033	ACCACCGGGCTGCTATCGGTATGGCAGACTGGATTATTGCTGCTCGAGTGGCTATT 139974
QY 380	gaagggtgctatcaattcggacaggatattccgagccagttgctaaaaatgccctttctc 439
Db 139973	GGAGCTCTTCTTCTTCGGGGGGGATGCGTCCATTGCACTCCGACAGCTCCCGGCC 139914
QY 440	aaccacaaacgaccgatcattggtcaatggcgcgagagaaagtgatgctcgtgtc 499
Db 139913	GGCCGTAAGTTGTCGTAGCGATCGCCAAACTCGTCCCAAGTGGCCNCTCATCGCTCTG 139854
QY 500	tatatgaatccatggtgccccgcggttga 531
Db 139853	GTACTCGCTCATCTCTCGCCCGTCGAGGTCA 139822

Search completed: September 9, 2000, 21:26:31
Job time: 4169 sec




```
Db 181 ACGTCGAGAAATACATATACAAATGGTTGGAAGAAATCCATTATGGCCAAACGCGGTGG 240
Qy 241 tgccggggtaaacggtaacgcatcacctgacggaagaaatgcaaaagagtttcatta 300
Db 241 TGCCGGGGTAAACCGGTAAACGCATCACCTGACGGAAGAAATGCAAAAGAGTTTCATTA 300
Qy 301 caccattggcccttattccaaaccgctctctgacatcgaaaccggtgaccggattattgt 360
Db 301 CACCATTGGCCCTTATTCCACACCGCTCTGACCATCGAACCCTGACCGGTGACCGGATTATTGT 360
Qy 361 cgacactcgagatgcttttgaagtgctatcaattcgaacagagatatccgagccatt 420
Db 361 CGACACTCGAGATGCTTTGAAGTGCATCAATTCGGAACAGGATATTCGAGCCAGTT 420
Qy 421 gctaaaaatgccctttctcaaccccaaaaacggaccgatcatggccaatggcgcgagaa 480
Db 421 GCTAAAAATGCCCTTCTCAACCCCAAAAACGGACCGATCATGTGTCATGGCGGAGAA 480
Qy 481 aggtgatgtgctgctgtctatatcgatccatgcttgcgcgcgcggttgatccctacgg 540
Db 481 AGGTGATGTGCTGCTGTCTATATCGAATCCATGTGTGCCGCGGTGATCCCTACGG 540
Qy 541 catctgcgccatgattccgcatctttggcgactgacccgggacccacctgacgccaatct 600
Db 541 CATCTGCCCATGATTCGCGATTTTGGCGGACTGACCCGGACCGACCTGACGGCGATGCT 600
Qy 601 caatgacccgtgcccagaaaaggctgcgactgattaaactcgacagtgaaaaggctactg 660
Db 601 CAATGATCCGTGCCAGAAAAGGTGCGCATGATTAAATCGACAGTGAAGGCTACTG 660
Qy 661 gagcaaacgcatacagcttccctataaaccataatgaccccttgagctatgcgcaga 720
Db 661 GAGCAAAACGCATAGCTTCCCTATAAACCCCATATTGGCACCTTGACGCTATCGCCAGA 720
Qy 721 aattgactcaatcaattcactgaqccagacaataccagcgcggaatatggatggcgga 780
Db 721 AATTGACTCAATCAATTCACTGACGCCAGACAATCAGCGCGGATATGGATGCGCGGA 780
Qy 781 tataggaccagggagattacactatcgcggtgaactgcgcgtgagcgccctgtttat 840
Db 781 TATAGGACACGAGGATATTACCTATCTGCCGCTACGTCCGCTGGAGCGCCCTGTTTAT 840
Qy 841 tggtagtcccatctgtcaggtgatgtgagatttgaggagccgagtagatttgc 900
Db 841 TGGTAGTCCCATCTGCTCAGGTGATGGTGAGATTGCGGGACCGCAGTAGAGTTGC 900
Qy 901 ctcaatcaccaccatcaaatgcattgatcaagaactggcagctttctctggccacgaat 960
Db 901 CTCAATCACCACCATCAAAAGTCGATTGATCAAGAAGTGGCAGCTTCTGCGCCACGAAT 960
Qy 961 ggagaatgcccgaataattatgatttggcagtgacgtgcgctgaggtgagatgcgacgcg 1020
Db 961 GGAGAATGCCGAAATATTATGATTTATGGCAGTGCACGTCCGCTGGAGGATGGACGCG 1020
Qy 1021 aattgcataatcgcaacttaattactggctggtagaagactttggttcgaacaatggga 1080
Db 1021 AATTGCATATCGCCACTTAATTACTGCTGGTAGAGACTTTTGGCTTCGAACAATGGGA 1080
Qy 1081 tgcctacatgcttctgagtcgaatgcgcaaatgctggcgcaacatggtcgaccacca 1140
Db 1081 TGCTACATGCTTCTGAGTCAATCGCGCAAGTGCAGCTGGGCAACATGCTCGACCCCAA 1140
Qy 1141 atacacgcttgccgcatgctgaacaaaaccctgttagtttagtagaataaactaacgg 1200
Db 1141 ATACACCGTTGGCGCATGTGAACAAAAACCTGTTAGTTTAGTAGAATAAATAAACCAG 1200
Qy 1201 tgaacattaccggatgtagatcgggtaattgtgaagtccaacacatcgctatttttaa 1260
Db 1201 TGAACATTACCGGATGATCGGGTAATGTGTAAGTTCAACAATCGCTATTTTTAA 1260
Qy 1261 cagctaaagcaggatgcataatggggccagatacaccatcaatatgtttactttactcc 1320
Db 1261 CAGCTAAAGCAGGTGCATATGGGGCCAGATACCCCATCAATATATGTTTACTTTACTCC 1320
```

```
Qy 1321 ttcaqcgagtgacggcgacacaaagttgtcacaatggcggaagcaaccaggtatt 1380
Db 1321 TTCAGCGAGTGACGGCGGACACAGAGTTGTACAAATGGCGGAGCAACCCAGGCTATT 1380
Qy 1381 gccaaaaatcaatcaaaatggcgcatcaacgacagaccactcaatgcaattcatttggat 1440
Db 1381 GCCAAATTAATCAAAATGGCGCATCAACGGCAGACCACTCAATGCAATTCATTGGAT 1440
Qy 1441 cc 1442
Db 1441 CC 1442
```

RESULT 2

Q66950/C
ID Q66950 standard; cDNA; 3900 BP.

AC Q66950;
DT 15-FEB-1995 (first entry)
DE Cell death hid cDNA.
KW Cell death; head involution defective; hid gene; reaper gene; rpr;
KW apoptosis; transgenic animal; antisense; ss.
OS Drosophila melanogaster.
FH Key Location/Qualifiers
FT cds 400..1632
FT /*tag= a

WO9416071-A.
PN 21-JUL-1994.
PD 14-JAN-1994; U00500.
PF 15-JAN-1993; US-004957.
PR 17-SEP-1993; US-123343.
PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
PI Abrams JM, Grether ME, Steller H, White K;
DR WPI: 94-249218/30.
DR P-PSDB: R55791

PT Isolated cell death genes from Drosophila - and novel assays for
PT apoptotic cell deaths and apoptotic and necrotic cell deaths,
PT based on selective staining by toluidine blue, acridine orange
PT and Nile blue
PS Disclosure: Fig. 6; 93pp; English.
CC The reaper (rpr) and head involution defective (hid) genes, mapping
CC to position 75C1,2 of D. melanogaster chromosome 3, exhibit
CC expression patterns related to the pattern of cell death during
CC Drosophila embryogenesis. Cell death genes, or antisense sequences,
CC can be used to reduce or abolish apoptosis, e.g. in transgenic
CC animals. An hid cDNA sequence is given in Q66950, and the
CC deduced sequence of the encoded HID protein in R55791.
SQ Sequence 3900 BP; 1288 A; 876 C; 736 G; 1000 T;

Query Match 2.7%; Score 39.2; DB 1; Length 3900;
Best Local Similarity 49.1%; Pred. No. 0.0077;
Matches 104; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

```
Qy 320 acaccctcctgaccatcgaaaccggtgaccgattattgtcgacactcgagatgctttt 379
Db 1458 ACCACCAGCTGGCTGCTGATGGCAGACTGGATTATTGCTGCTGCTGAGTGGCTATT 1399
Qy 380 gaagggtcatcaattcgaacagagatatccgagccagttgctaaaaatgccctttctc 439
Db 1398 GGAGCTCTTTCTTCTCCGGGGGGGATGCTGCTCAATTTGAACCTCTGACAGCTCCCGGCC 1339
Qy 440 aaccaccaaaacgaccgatcatggttcaatggcgcgagaaaggatgctcgctgtc 499
Db 1338 GGCCGTAAGTTTGTGCTAGCGATCGCAAACTCGTCCCAAGTGGCGTCAATGATCGCTCTG 1279
Qy 500 tatatgaatccatgttggcccgcggttga 531
Db 1278 GTACTCGCGCTCATCTCTCGCCCGTCGAGGTCA 1247
```

RESULT 3

Q66948/C

```
ID Q66948 standard; DNA; 502 BP.
AC Q66948;
AT 15-FEB-1995 (first entry)
DE Cell death hid gene.
KW Cell death; head involution defective; hid gene; reaper gene; rpr;
OS apoptosis; transgenic animal; antisense; ss.
FH Drosophila melanogaster.
FS Key Location/Qualifiers
FT misc_difference 93
FT /*tag= a
FT /note= "Base n at position 93 is not identified
FT in the specification"
FN W09416071-A.
PD 21-JUL-1994.
PF 14-JAN-1994; U00500.
PR 15-JAN-1993; US-004957.
PR 17-SEP-1993; US-123343.
PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.
PI Abrams JM, Grether ME, Steller H, White K;
DR WPI; 94-249218/30.
DR P-PSDB; R5790.
PT Isolated cell death genes from Drosophila - and novel assays for
PT apoptotic cell deaths and apoptotic and necrotic cell deaths,
PT based on selective staining by toluidine blue, acridine orange
PT and Nile blue
PS Disclosure; Fig.5; 93pp; English.
CC The reaper (rpr) and head involution defective (hid) genes, mapping
CC to position 75C1.2 of D. melanogaster chromosome 3, exhibit
CC expression patterns related to the pattern of cell death during
CC Drosophila embryogenesis. Cell death genes, or antisense sequences,
CC can be used to reduce or abolish apoptosis, e.g. in transgenic
CC animals. The genomic sequence of hid is given as 4 contiguous
CC sequences in Q66946-49, which together encode HID protein.
CC Sequence 502 BP; 141 A; 135 C; 135 G; 90 T;
SQ
```

```
Query Match 2.6%; Score 37.6; DB 1; Length 502;
Best Local Similarity 48.6%; Pred. No. 0.0072;
Matches 103; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

Oy 320 acaccgccctgaccatcgaaacccggtgaccggtattgtgcacacgcagatgctttt 379
|| || || || || || || || || || || || || || || || || || || || ||
Db 333 ACCACCGACATGGCTGTCGGTATGCGACACTGGATTATTCGTCGTCGAGTGGCTATT 274

Oy 380 gaagtgctatcaattcggaacagatattccgagccagtggtctaaatgccctttctc 439
|| || || || || || || || || || || || || || || || || || || || ||
Db 273 GGAGCTCTCTCTTCCTCGGGGGGGGATGCGTCCATTGAACCTCTCGACACGCTCCGGCC 214

Oy 440 aaccacaaaacgaccgatcatgtcgaatggcgcggaaggtgatgtcgtcgtctgc 499
|| || || || || || || || || || || || || || || || || || || || ||
Db 213 GGCCGTAAGTTGTGCTAGCGATCGCAAACTCGTCCCAAGTGGCTCATGATCGCTCTG 154

Oy 500 tatatcgaatccatgttgcggcgccgcttga 531
|| || || || || || || || || || || || || || || || || || || || ||
Db 153 GTACTCGGCTCATCTCGCCCGTCGAGGTCA 122
```

```
RESULT 4
X28300/C
ID X28300 standard; DNA; 429 BP.
AC X28300;
AT 17-JUN-1999 (first entry)
DE Human CYP3A4 gene exon 7, intron 7.
KW CYP3A4 gene polymorphism; polymorphic locus; human; altered metabolism;
KW CYP3A4 substrate; drug-drug interaction identification; toxin exposure;
KW genetic linkage detection; phenotypic variation; intron; ss.
OS Homo sapiens.
PN W09913106-A1.
PD 18-MAR-1999.
PF 02-SEP-1998; U18158.
PR 10-SEP-1997; US-058612.
PA (AXIS-) AXIS PHARM INC.
PI Guida M, Lichter JB;
```

```
DR WPI; 99-215070/18.
PT New isolated CYP3A4 polymorphic sequences
PS Disclosure; Page 30; 40pp; English.
CC This sequence represents an intron of the human CYP3A4 gene.
CC The invention relates to a CYP3A4 sequence polymorphism,
CC which is part of a non-naturally occurring chromosome. Nucleic acids
CC comprising the CYP3A4 polymorphic sequences can be used to screen
CC patients for altered metabolism for CYP3A4 substrates, potential
CC drug-drug interactions, and adverse/side effects as well as diseases that
CC result from environmental or occupational exposure to toxins. They can
CC also be used to establish animal, cell culture and in vitro cell-free
CC models for drug metabolism. Polymorphic CYP3A4 gene sequences can be used
CC for expression studies to determine the effect of promoter and/or intron
CC sequence variations on mRNA expression and stability. The polymorphisms
CC are also used as single nucleotide polymorphisms to detect genetic
CC linkage to phenotypic variation in activity and expression of CYP3A4. The
CC nucleic acids can also be used to generate genetically modified non-human
CC animals or site specific gene modifications in cell lines.
SQ Sequence 429 BP; 126 A; 95 C; 53 G; 153 T;

Query Match 2.4%; Score 35.2; DB 1; Length 429;
Best Local Similarity 57.5%; Pred. No. 0.044;
Matches 61; Conservative 1; Mismatches 44; Indels 0; Gaps 0;

Oy 1164 acaaaaaccgttagtttagtaggaataactaacccggtgaacattaccggatgtagtc 1223
|||| ||| || || || || || || || || || || || || || || || || || || ||
Db 315 ACACATATCTTCCAAATGTACTACAAATCACTGAACCTGTATATTTAAGTGGATGATTAC 256

Oy 1224 ggggtaagtgtgaagtccaacaatcgctattttaacagctaaag 1269
|||| ||| || || || || || || || || || || || || || || || || || || ||
Db 255 ATGGTGATTATATCTCAATAAAGCAGTTATTTTAAAGAGAGAGMAAG 210
```

```
RESULT 5
X00477
ID X00477 standard; DNA; 5733 BP.
AC X00477;
AT 30-MAR-1999 (first entry)
DE Arabidopsis thaliana clavatal encoding DNA.
KW Arabidopsis thaliana; clavatal; CIV1; meristem phenotype; harvesting;
KW apical shoot enlargement; floral meristem; resistance; ss.
OS Arabidopsis thaliana.
FH Key Location/Qualifiers
FT CDS 2434..5470
FT /*tag= a
FT /note= "contains introns"
FT exon 2434..5038
FT /*tag= b
FT intron 5039..5117
FT /*tag= c
FT exon 5118..5470
FT /*tag= d
FT /*tag= 2
FT US5859338-A.
PN 12-JAN-1999.
PD 06-JUN-1995; 473553.
PR 06-JUN-1995; US-473553.
PA (CALY ) CALIFORNIA INST OF TECHNOLOGY.
PI Clark SE, Meyerowitz EM, Williams RW;
DR WPI; 99-120031/10.
DR P-PSDB; W30607.
PT Isolated nucleic acid encoding the plant clavatal protein and
PT related vectors, plant cells and plants - used to regulate meristem
PT phenotype, for increasing yield of leaves, flowers, fruits and seeds
PS Claim 3; Fig 5; 45pp; English.
CC The present sequence encodes clavatal (CIV1) isolated from Arabidopsis
CC thaliana. The nucleotide sequence encoding CIV1, including its
CC corresponding antisense sequences, are used to alter meristem phenotypes,
CC in particular the enlargement of apical shoot or floral meristems, to
CC provide an increased yield of leaves, flowers, fruits and seeds in a
```

CC very wide variety of plants. They may also be used to increase resistance
 CC to lodging or wind damage, or to facilitate harvesting. Fragments of the
 CC nucleotide sequence encoding CnV1 are also useful as probes to identify
 CC related sequences in other plants.
 SQ Sequence 5733 BP; 1732 A; 1144 C; 1094 G; 1763 T;

Query Match 2.4%; Score 34.2; DB 1; Length 5733;
 Best Local Similarity 50.3%; Pred. No. 0.54;
 Matches 84; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

QY 1136 cccaatacaccttgccgcatgctgaacaaaacctgttagtttagtaggaataaacta 1195
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 3709 CTCACGGCACCTGTTCCGCGGGGCTTTCAATCTACCGTTAGTACGATTATCGAACTC 3768
 QY 1196 accggtgaacattaccgcgatgtagatcggtgtaagttaagtccaacaaactcgtatt 1255
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 3769 ACTGATAATTTCTTCGCGTGAACCTCGGTAACGATGTCGGCGGATGTTCTCGATCAG 3828
 QY 1256 tttaacagcgttaacgaggtgcataatggtggccagatcacacccatcaat 1302
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 3829 ATTTACCTCTCTAACAACCTGGTTTCCGGCGAGATTCCACCTGCCAT 3875

RESULT 6
 X14578
 ID X14578 standard; DNA; 745 BP.
 AC X14578;
 DT 31-MAR-1999 (first entry)
 DE H. pylori GHPO 1662 gene.
 KW GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
 KW peptic ulcer disease; ss.
 OS Helicobacter pylori.
 FH Key Location/Qualifiers
 FT CDS 37..720
 FT /*tag= a
 PN W09843478-Al.
 PD 08-OCT-1998.
 PF 01-APR-1998; U06371.
 PR 29-JUL-1997; US-902615.
 PR 01-APR-1997; US-833457.
 PR 24-JUN-1997; US-881227.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA (INMR-) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
 PI Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;
 DR WPI; 98-542233/46.
 DR P-PSDB; W98859.
 CC New isolated Helicobacter polynucleotides - used to develop products
 CC for the diagnosis, prevention and treatment of Helicobacter
 CC infections and gastrointestinal diseases
 PS Claim 1; Page 1937-1938; 2054pp; English.
 CC This sequence represents a polynucleotide of the invention. It was
 CC isolated from Helicobacter pylori and encodes a H.pylori GHPO protein.
 CC The polypeptides can be used for preventing or treating Helicobacter
 CC infections, and gastroduodenal diseases associated with these
 CC infections, including acute, chronic, and atrophic gastritis, and peptic
 CC ulcer diseases, e.g. gastric and duodenal ulcers. They can also be used
 CC for the production of antibodies. The products can also be used for
 CC detection and diagnosis.
 SQ Sequence 745 BP; 265 A; 102 C; 159 G; 219 T;

Query Match 2.3%; Score 33.2; DB 1; Length 745;
 Best Local Similarity 48.9%; Pred. No. 0.32;
 Matches 89; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

QY 176 aaattacgtgagagaataacatatgaatggttggaagaatccattatggccaaacgcggt 235
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 31 AAATTAATGGACAGAGAGATGAATAAATCAACAAATAGTCATTATCAATCGGCC 90
 QY 236 gttggtccgggcgttaacacggttaacgcatcacctgcgaggaagaataacaaagagttt 295
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 91 ATTCAGGTGGGGCAAGACCTCTTTGATCAACACAGATTGAGATGGCAAAAGCTTG 150

QY 296 cattacacacattggcccttatttccacacccgctcctgacacatgaacccgggtgaccgatt 355
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 151 GGGCATTCTATTAGCGTTCAITCTTACCGATGAATATTTCCAAACACAGATGAAGAGGT 210
 QY 356 at 357
 ||
 Db 211 AT 212

RESULT 7
 Q89701/c
 ID Q89701 standard; DNA; 3201 BP.
 AC Q89701;
 DT 27-DEC-1995 (revised)
 DT 31-OCT-1995 (first entry)
 DE Alpha-1,4-glucan lyase gene.
 KW Alpha-1,4-glucan lyase; enzyme; ss.
 OS Morchella costata.
 FH Key Location/Qualifiers
 FT CDS 1..3201
 FT /*tag= a
 PN W09510617-A.
 PD 20-APR-1995.
 PF 15-OCT-1994; E03398.
 PR 15-OCT-1993; GB-021302.
 PA (DANI-) DANISCO AS.
 PI Boisen K, Christensen TMIE, Kragh KM, Marcussen J;
 PI Yu S;
 DR WPI; 95-161802/21.
 DR P-PSDB; R72711.
 DR Isolation of alpha-1,4-glucan lyase from fungus - and its
 PT recombinant prodn from isolated DNA.
 PS Claim 9; Pages 30-32; 51pp; English.
 CC The enzyme alpha-1,4-glucan lyase(GL) was purified from Morchella
 CC costata (Mc) by affinity chromatography on beta-cyclodextrin
 CC Sepharose (RTM), ion exchange on Mono Q HR 5/5 (RTM) and gel
 CC filtration on Sepharose (RTM) 12 columns. The purified enzyme
 CC appeared colourless. The mol. mass was 110 kDa (SDS-PAGE). It showed
 CC an isoelectric point of pI 5.4. The optimum pH range for the fungal
 CC lyase catalysed reaction was between pH 5 and pH 7. The purified
 CC fungal lyase degraded maltosaccharides. The highest activity was
 CC with maltotetraose (activity 100%), then maltohexaose (97%),
 CC maltotriose (56%) and maltose (2%). The lyase was digested with
 CC endoproteinase Arg-C from Clostridium histolyticum or endoproteinase
 CC Lys-C from Lysobacter enzymogenes. The resulting peptides were
 CC sequenced (see FT). Amino acid sequences of three overlapping
 CC tryptic peptides from GL (see R72711) were used to generate mixed
 CC oligos which could be used as PCR primers for the amplification of
 CC DNA isolated from Morchella. The primers are Q89703-4 and Q90305-6.
 SQ Sequence 3201 BP; 903 A; 752 C; 770 G; 776 T;

Query Match 2.3%; Score 33.2; DB 1; Length 3201;
 Best Local Similarity 50.0%; Pred. No. 0.83;
 Matches 83; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

QY 243 ccgggcgttaaacgcgttaacgcattcacctacgcggaagaataacaaagatttcattaca 302
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 1028 CTGGGCTTCAACTTTGACGTCCTCAACACACCTGTATAGTCGTACAATTTCCGGGACC 969
 QY 303 ccattggcccttattccacacccgctcctgacacatcgaaacccggtgaccggattattgtcg 362
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 968 GTATCGCACTGATACCGTAACAAATCGATACCACTACTCTGGTTCCTCCAGCTTGATGTAG 909
 QY 363 acactcgagatgcttttgaagggtgctatcaattcggaaacaggatat 408
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 908 CTGAGTTGTTGTTCTTCCAAAGTCGTGGCAATTTGAGAGTAGTATAT 863

RESULT 8
 Q87616/c
 ID Q87616 standard; DNA; 3201 BP.

AC Q87616;
 DT 10-JAN-1996 (first entry)
 DE M.costata DNA encoding alpha-1,4-glucan lyase.
 KW Alpha-1,4-glucan lyase; 1,5-D-anhydrofructose; G.lemaleiformis;
 KW antioxidant; food additive; sweetener; ds.
 OS Morchella costata.
 FH key Location/Qualifiers
 FT 1..3201
 FT cds
 FT /*tag= a
 FT /product= alpha_1,4-glucan_lyase
 PN W09510616-A2.
 PD 20-APR-1995.
 PF 15-OCT-1994; E03397.
 PR 15-OCT-1993; GB-021302.
 PR 15-OCT-1993; GB-021301.
 PR 15-OCT-1993; GB-021303.
 PR 15-OCT-1993; GB-021305.
 PR 15-OCT-1993; GB-021304.
 PR (DANI-) DANISCO AS.
 PA Bojko M, Bojzen K, Christensen TMIE, Kragh KM, Marcussen J;
 PI Nielsen J, Yu S;
 DR WPI: 95-161801/21.
 DR P-PSDB: R70638.
 DR 1,5-D-anhydrofructose prodn. from alpha-1,4-glucan - by treatment
 PT with pure alpha-1,4-glucan lyase, partic. useful as antioxidant and
 PT sweetener for foods and beverages
 PS Claim 12; Page 101c-102b; 166pp; English.
 CC An alpha-1,4-glucan lyase enzyme (preferably pullanase or isoamylase)
 CC is used in a new method for the production of 1,5-D-anhydrofructose.
 CC The enzyme is isolated from either a fungus (M.costata or M.vulgaris)
 CC or from fungally infected algae (G.lemaleiformis) or algae alone.
 CC 1,5-D-anhydrofructose is useful as an antioxidant and sweetener for
 CC foodstuffs and beverages. It is also useful as an intermediate for the
 CC antibiotic microthecin, an oxygen scavenger during polymerisation
 CC reactions and as a reducing agent in the synthesis of biodegradable
 CC plastics.
 SQ Sequence 3201 BP; 903 A; 752 C; 770 G; 776 T;

.Query Match 2.3%; Score 33.2; DB 1; Length 3201;
 Best Local Similarity 50.0%; Pred. No. 0.83;
 Matches 83; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

QY 243 ccgggctgaacggtaacgcatcacctgacggaagaatgcacaaagagtttcattaca 302
 Db 1028 CTGGGCTTCAACCTTGACGCTCAACAGACCTGTATAAAGTCGTACAAATTTCCGGGACC 969
 QY 303 ccattggcccttattccacacccgctctgacccatgcacccggtgacccgattattgtcg 362
 Db 968 GTATCCGCACTGATACCGTAACAAATCGATACCACTACCTGTTCCACAGCTTGATGTAG 909
 QY 363 acactcgagatgctttgaagtgctatcaattcggaacagatat 408
 Db 908 CCTGAGTTGGCTTTCCAAAGTCGATGGCAATTTGAGTAGTTAT 863

RESULT 9
 V84194/c
 ID V84194 standard; DNA; 3201 BP.
 AC V84194;
 DT 29-MAR-1999 (first entry)
 DE Morchella costata glucan lyase DNA.
 KW Glucan lyase; antioxidant; transgenic plant; stress tolerance;
 KW anhydrofructose; beverage; wine; foodstuff; ds.
 OS Morchella costata.
 PN W09850532-A2.
 PD 12-NOV-1998.
 PF 06-MAY-1998; IFB0708.
 PR 06-MAY-1997; GB-009161.
 PA (DANI-) DANISCO AS.
 PI Buchter-Larsen A, Marcussen I;
 DR WPI: 99-070094/06.
 DR P-PSDB: W88255.

PT Recombinant production of anti-oxidant compounds - by the production
 PT of an anhydrofructose from a glucan, used for improving plants for
 PT use in foodstuffs
 PS Claim 9; Page 45-46; 53pp; English.
 CC This nucleotide sequence codes for a glucan lyase (see W88255) of
 CC Morchella costata used in claimed processes of the invention. A
 CC claimed process of preparing a medium that contains an antioxidant
 CC and at least one other component involves preparing antioxidant
 CC from a glucan, in situ, by use of recombinant DNA techniques. Also
 CC claimed are: (1) use of anhydrofructose as an antioxidant for a
 CC medium, where the anhydrofructose is prepared in situ in the
 CC medium; (2) use of anhydrofructose for imparting or improving
 CC stress tolerance in a plant, where the anhydrofructose is prepared
 CC in situ in the plant; (3) use of glucan lyase for imparting or
 CC improving: (i) stress tolerance in plant, or (ii) transformation of
 CC grape, where the glucan lyase is prepared in situ; (4) use of a
 CC nucleotide sequence (NS) coding for a glucan lyase as a means of
 CC imparting or improving stress tolerance in a plant, where the NS is
 CC expressed in situ in the plant, and (5) use of a NS coding for a
 CC glucan lyase for imparting or improving the transformation of a
 CC grape, where the NS is expressed in situ in the grape. The
 CC antioxidant containing medium can be used as foodstuff or in the
 CC preparation of foodstuffs such as beverages, in particularly
 CC alcoholic beverages such as wine (claimed).
 SQ Sequence 3201 BP; 903 A; 752 C; 770 G; 776 T;

Query Match 2.3%; Score 33.2; DB 1; Length 3201;
 Best Local Similarity 50.0%; Pred. No. 0.83;
 Matches 83; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

QY 243 ccgggctgaacggtaacgcatcacctgacggaagaatgcacaaagagtttcattaca 302
 Db 1028 CTGGGCTTCAACCTTGACGCTCAACAGACCTGTATAAAGTCGTACAAATTTCCGGGACC 969
 QY 303 ccattggcccttattccacacccgctctgacccatgcacccggtgacccgattattgtcg 362
 Db 968 GTATCCGCACTGATACCGTAACAAATCGATACCACTACCTGTTCCACAGCTTGATGTAG 909
 QY 363 acactcgagatgctttgaagtgctatcaattcggaacagatat 408
 Db 908 CCTGAGTTGGCTTTCCAAAGTCGATGGCAATTTGAGTAGTTAT 863

RESULT 10
 V88779
 ID V88779 standard; cDNA; 383 BP.
 AC V88779;
 DT 12-FEB-1999 (first entry)
 DE EST clone H0266.
 KW Expressed sequence tag; secreted protein; haematopoiesis regulator;
 KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
 KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
 KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
 OS Homo sapiens.
 PN W09845437-A2.
 PD 15-OCT-1998.
 PF 10-APR-1998; U06956.
 PR 10-APR-1997; US-837312.
 PA (GEMT) GENETICS INST INC.
 PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,
 PI Racie LA, Spaulding V, Treacy M;
 DR WPI: 99-070078/06.
 PT New polynucleotides encoding human secreted proteins - derived from
 PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
 PT ovary, pituitary, retina and colon cDNA libraries
 PS Claim 1; Page 509-510; 641pp; English.
 CC The present sequence represents an expressed sequence tag (EST), and is
 CC a polynucleotide of the invention. The polynucleotides of the invention
 CC are all secreted EST sequences isolated from a variety of human tissue
 CC sources. The EST sequences and proteins encoded by them are predicted to
 CC have useful biological activities which would make them suitable for
 CC treating, preventing or ameliorating medical conditions in humans and

CC animals, although no supporting data is given. Suggested activities
CC include nutritional activity, immune stimulating or suppressing activity,
CC haematopoiesis regulating activity, tissue growth activity,
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition
CC activity. The EST sequences are also stated to be useful for gene
CC therapy. 383 BP: 92 A: 108 C: 117 G: 66 T:
SQ Sequence

Query Match 2.3%; Score 32.8; DB 1; Length 383;
Best Local Similarity 50.0%; Pred. NO. 0.28;
Matches 82; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

Qy	535		ctacggcatgcccgaatgatccgcgaattttgcgactgacggagaccgacctgacgcg	594
Dδ	194		CAGGACTTCTTCAGCATGAAGCCGGAGTGCGAGAATTGTAAACCACTCCAACGTGGCGG	253
Qy	595		catagtcaaatgatccgctgccagaanaaggtcgcgatattaaactgcagtgaanaagt	654
Dδ	254		CATGCACACGCCGTGGGCTAACGAGTTCATCGTAGAAMAATCCGGGAGCACAAGT	313
Qy	655		ctactggagcaaacgcacatcgttccctataaaaacccccatttg	698
Dδ	314		TGTTTTTGCTCAACATGCTGGGCTCCCGCACACCAGCAATGTGTG	357

RESULT	11
	Q57702/c
ID	Q57702 standard; cDNA; 1226 BP.
AC	Q57702;
DT	15-AUG-1994 (first entry)
DE	K lactis transaldolase gene KlTAL1 portable promoter.
KW	Kluyveromyces lactis; budding yeast; transaldolase; strong promoter;
KW	portable promoter; KlTAL1; ds.
OS	Kluyveromyces lactis.
PN	W09403618-A.
PD	17-FEB-1994.
PF	28-JUL-1993; F00771.

PF	28-JUL-1993; F00771.
PR	30-JUL-1992; FR-009432.
PA	(RHON) RHONE POULENC RORER SA.
PI	Bolotin M., Menart S;
DI	WPI; 94-065706/08.
DR	
PT	New transcriptional promoter from the Kluyveromyces lactis
PT	trans-aldolase gene - providing high level expression of
PT	heterologous protein, esp. human serum albumin, in yeast, not
PT	subject to glucose repression
PT	Claim 3; Page 16; 30pp; French.
PS	
CC	A 1.3kb cDNA fragment (057701) was isolated from a library of
CC	K.lactis 2359/152 genome fragments fused to E.coli lacZ. The
CC	isolated clone comprises a strong promoter sequence suitable for
CC	high level expression of recombinant proteins in Kluyveromyces or
CC	other yeast hosts. The promoter is not subject to glucose repression
CC	so can be used in conventional culture media. Restriction enzyme
CC	sites were introduced into the 1.3kb fragment by PCR. The PCR product
CC	was cloned in vector pCRII to give pYG176. The KLTAL1 promoter can be
CC	removed from this plasmid by MluI-HindIII digestion and the
CC	resulting "portable promoter" is then used in the construction of
CC	vectors for expression of heterologous proteins.
CC	Sequence 1226 BP; 356 A; 241 C; 188 G;
SQ	441 T;

Query Match 2.2%; Score 32.2; DB 1; Length 1226;
Best Local Similarity 63.6%; Pred. No. 0.98;
Matches 49; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

Qy	1160	ctgaacaaacac	ctgttagttag	taggaataac	taacccggtga	cattacccg	gatgta	1219
Db	568	CGGATGGACA	CCGGGTAGT	TATCCGGGAT	CACTTCCAGT	GATCCTT	GCCGTGGT	CGTA 509
Qy	1220	gatacggggta	atgtgtgta	1236				

Db 508 GATGGAGCTATATTTTA 492

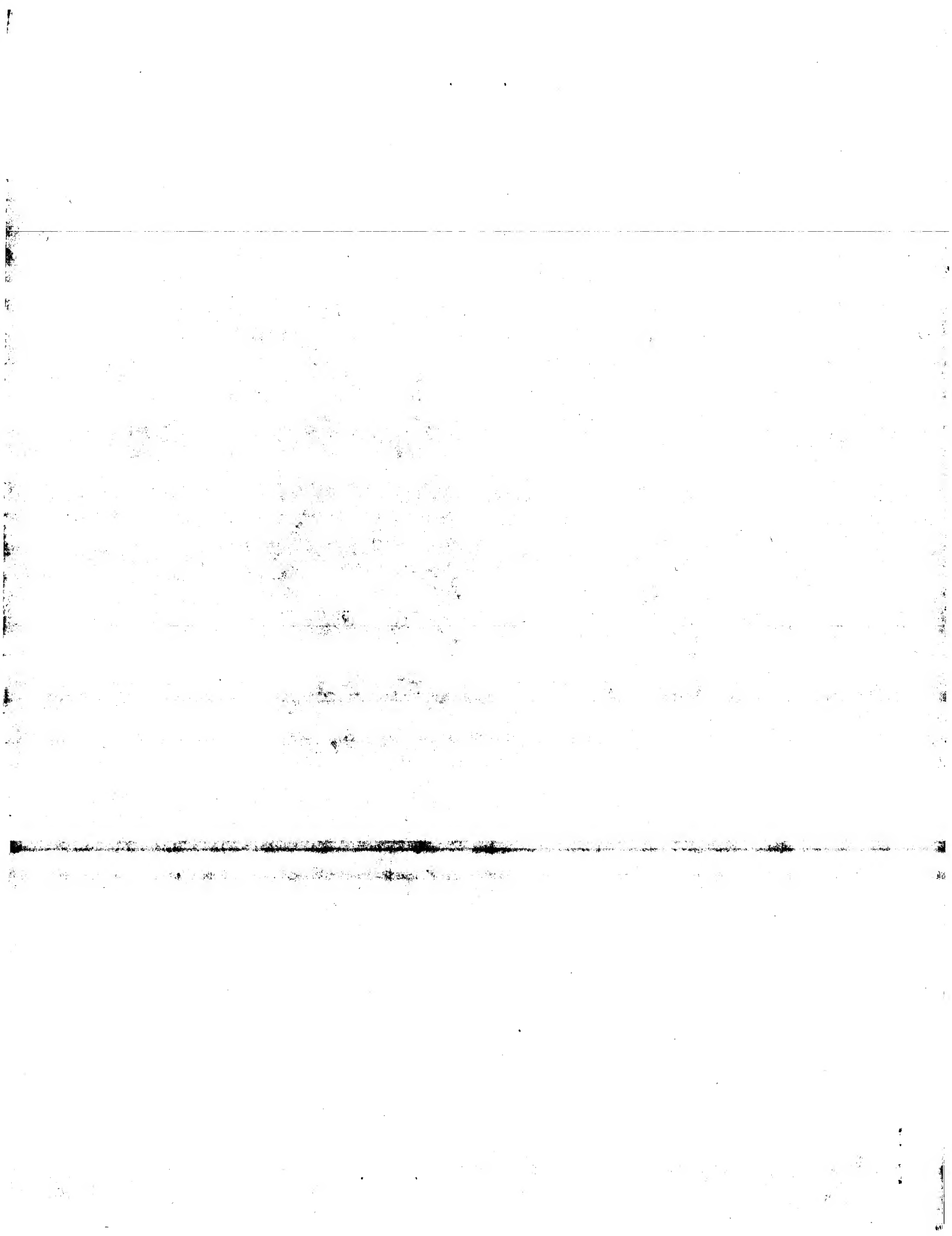
```
RESULT 12
Q57701/C
ID Q57701 standard; cDNA; 1349 BP.
AC Q57701;
DT 15-AUG-1994 (first entry)
DE K.lactis transaldolase gene KITAL1 promoter.
KW Kluyveromyces lactis; budding yeast; transaldolase; strong promoter;
KW portable promoter; KITAL1; ds.
OS Kluyveromyces lactis.
FH key location/Qualifiers
FT promoter 1. .1296
FT FT /*tag= a
FT cds 1297. .1349
FT FT /*tag= b
FT FT /product= transaldolase_N-terminal
FT FT /note= "partial CDS"
FN W09403618-A.
FD 17-FEB-1994.
PF 28-JUL-1993; F00771.
PR 30-JUL-1992; FR-009432.
PA (RHON ) RHONE POULENC RORER SA.
PI Bolotin M, Menart S;
DR WPI; 94-065706/08.
DR P-PSDS: R46601.
PT New transcriptional promoter from the Kluyveromyces lactis
PT trans-aldolase gene - providing high level expression of
PT heterologous protein, esp. human serum albumin, in yeast, not
PT subject to glucose repression
PS Claim 1; Page 14-15; 30pp; French.
CC A 1.3kb cDNA fragment (Q57701) was isolated from a library of
CC K.lactis 2359/152 genome fragments fused to E.coli lacZ. The
CC isolated clone comprises a strong promoter sequence suitable for
CC high level expression of recombinant proteins in Kluyveromyces or
CC other yeast hosts. The promoter is not subject to glucose repression
CC so can be used in conventional culture media. See also Q57702.
SQ sequence 1349 BP; 397 A; 271 C; 202 G; 479 T;
```

	Query Match	2.2%	Score 32.2;	DB 1;	Length 1349;
	Best Local Similarity	63.6%;	Pred. No. 1;		
	Matches 49;	Conservative 0;	Mismatches 28;	Indels 0;	Gaps 0;
Qy 1160	ctgaacaaacacctgttagtttagtagaataacacccgtgaacattaccgcgatgta	1219			
Db 640	CGGATGGACAAACCGGTAGTTTATTCGGGGATGACTTTCACGTGATCCTTGCTGGTCGTA	581			
Qy 1220	gatcgggggtaatgtgta	1236			
Db 580	GATGGAGCTATATTTTA	564			

RESULT	13
V52729	ID V52729 standard; DNA; 6260 BP.
V52729	AC V52729;
DT	21-DEC-1998 (first entry)
DE	Human hepatocyte nuclear factor 1 beta gene (TCF2).
KW	Hepatocyte nuclear factor 1 beta; HNF-1 beta; MODY4; human;
KW	transcription factor; maturity onset diabetes of the young;;
KW	diabetes; NIDDM; diagnosis; therapy; ss.
OS	Homo sapiens.
FH	Key
FT	protein_bind
FT	14..31
FT	Location/Qualifiers
FT	/*tag= a
FT	/label= NF-1
FT	34..46
FT	/*tag= b
FT	/label= HNF-3
FT	59..69
FT	protein_bind
FT	/*tag= c
FT	
FT	

FT /label= AP1
 FT 113. .131
 FT /*tag= d
 FT /label= NF-1
 FT 154. .163
 FT /*tag= e
 FT /label= AP4
 FT 171. .180
 FT /*tag= f
 FT /label= AP4
 FT 201. .212
 FT /*tag= g
 FT /label= HNF-3
 FT 284. .300
 FT /*tag= h
 FT /label= NF-1
 FT 292. .300
 FT /*tag= i
 FT /label= AP4
 FT 307. .316
 FT /*tag= j
 FT /label= AP4
 FT 389. .400
 FT /*tag= k
 FT /label= AP4
 FT 421. .450
 FT /*tag= l
 FT /label= HNF-4alpha
 FT 459. .470
 FT /*tag= m
 FT /label= HNF-3
 FT 479. .490
 FT /*tag= n
 FT /label= C/EBP
 FT 492. .514
 FT /*tag= o
 FT /label= HNF-3
 FT 516. .537
 FT /*tag= p
 FT /label= AP1
 FT 715. .6215
 FT /*tag= q
 FT /note= "contains introns"
 FT 490. .1040
 FT /*tag= r
 FT /number= 1
 FT 1041. .1756
 FT /*tag= s
 FT /number= 1
 FT misc_difference 1352
 FT /*tag= t
 FT /note= "n at position 1352 represents an intronic
 FT sequence of 9 kb"
 FT exon 1757. .1956
 FT /*tag= u
 FT /number= 2
 FT 1957. .2588
 FT /*tag= v
 FT /number= 2
 FT misc_difference 2205
 FT /*tag= w
 FT /note= "n at position 2205 represents an intronic
 FT sequence of 3.8 kb"
 FT exon 2589. .2752
 FT /*tag= x
 FT /number= 3
 FT 2753. .3210
 FT /*tag= y
 FT /number= 3
 FT 3211. .3452
 FT /*tag= z
 FT /number= 4
 FT 3453. .3739
 FT intron

FT /*tag= aa
 FT /number= 4
 FT misc_difference 3564
 FT /*tag= ab
 FT /note= "n at position 3564 represents an intronic
 FT sequence of 2 kb"
 FT exon 3740. .3901
 FT /*tag= ac
 FT /number= 5
 FT 3902. .4018
 FT /*tag= ad
 FT /number= 5
 FT 4019. .4220
 FT /*tag= ae
 FT /number= 6
 FT 4221. .4671
 FT /*tag= af
 FT /number= 6
 FT misc_difference 4339
 FT /*tag= ag
 FT /note= "n at position 4339 represents an intronic
 FT sequence of 0.8 kb"
 FT exon 4672. .4863
 FT /*tag= ah
 FT /number= 7
 FT 4864. .5326
 FT /*tag= ai
 FT /number= 7
 FT misc_difference 5177
 FT /*tag= aj
 FT /note= "n at position 5177 represents an intronic
 FT sequence of 1.5 kb"
 FT exon 5327. .5448
 FT /*tag= ak
 FT /number= 8
 FT 5449. .5541
 FT /*tag= al
 FT /number= 8
 FT 5542. .5686
 FT /*tag= am
 FT /number= 9
 FT 5687. .6087
 FT /*tag= an
 FT /number= 9
 FT misc_difference 5853
 FT /*tag= ao
 FT /note= "n at position 5853 represents an intronic
 FT sequence of 1.5 kb"
 FT exon 6088. .6260
 FT /*tag= ap
 FT /number= 10
 FT WO9811254-A1.
 FT 19-MAR-1998.
 FT 10-SEP-1997; U16037.
 FT 30-OCT-1996; US-029679.
 FT 10-SEP-1996; US-025719.
 FT 02-OCT-1996; US-028056.
 FT (ARCH-) ARCH DEV CORP.
 FT Bell GI, Furuta H, Horikawa Y, Kaisaki PJ, Menzel S,
 FT Oda N, Yamagata K;
 FT WPI; 98-271667/24.
 FT DR P-PSDB; W71580.
 FT Isolated nucleic acid encoding hepatocyte nuclear factor 1-alpha and
 FT 1-beta - useful for detecting susceptibility for non-insulin
 FT dependent diabetes, especially maturity-onset diabetes of the young
 FT Claim 95; Fig 26A-M; 363pp; English.
 FT PS This is a partial nucleotide sequence of the gene (TCF2) encoding
 FT CC human hepatocyte nuclear factor 1 beta (HNF-1 beta, see W71580), a
 FT CC homeodomain-containing transcription factor. Mutations in this
 FT CC gene are indicative of a propensity to diabetes mellitus. The
 FT CC invention concerns the identification of genes responsible for
 FT non-insulin dependent diabetes mellitus (NIDDM) for use in
 FT diagnostics and therapeutics. It demonstrates that the MODY3 locus



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 9, 2000, 20:16:07 ; Search time 52.2 Seconds
(without alignments)
3799.914 Million cell updates/sec

Title: US-09-214-679-1

Perfect score: 1442

Sequence: 1 cccgggaactccattgtgcc.....aatgcaattcattggtatcc 1442

Scoring table:

IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 243080 seqs, 68777915 residues

Total number of hits satisfying chosen parameters: 486160

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA.*

1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/5C_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/5D_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/6_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/PCUS_COMB.seq.*
7: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	37.6	2.6	503	1	US-08-123-343A-15
C 2	37.6	2.6	3900	1	US-08-123-343A-6
C 3	34.2	2.4	5733	3	US-08-473-553A-1
C 4	33.2	2.3	3201	3	US-08-633-770A-3
C 5	33.2	2.3	4726	3	US-08-633-770A-11
C 6	32.8	2.3	7218	1	US-08-232-463-14
C 7	32.4	2.2	9551	1	US-08-056-200-93
C 8	32.4	2.2	9551	4	US-08-800-644-93
C 9	32.2	2.2	1226	1	US-08-374-686-4
C 10	32.2	2.2	1349	1	US-08-374-686-1
C 11	31.2	2.2	2167	4	US-08-743-637B-16
C 12	31.2	2.2	2167	5	US-08-526-840B-16
C 13	30.8	2.1	1187	6	PCR-US95-13536-2
C 14	30.8	2.1	1188	6	PCR-US95-13536-1
C 15	30.2	2.1	10095	5	US-08-822-586-45
C 16	30	2.1	2665	5	US-09-040-005-1
C 17	29.4	2.0	1888	2	US-08-485-107-1
C 18	29.2	2.0	46899	2	US-08-471-119A-1
C 19	29	2.0	300	1	US-08-083-946-6
C 20	29	2.0	300	5	US-08-452-915-6
C 21	29	2.0	850	1	US-07-611-528A-1
C 22	29	2.0	850	1	US-08-083-946-1
C 23	29	2.0	850	5	US-08-452-915-1
C 24	28.6	2.0	1212	4	US-09-092-770-18
C 25	28.4	2.0	2380	7	5268463-1
C 26	28.4	2.0	2438	7	5432081-1

Query Match 2.6%; Score 37.6; DB 1; Length 503;

ALIGNMENTS

RESULT 1

US-08-123-343A-15/c

; Sequence 15, Application US/08123343A

; Patent No. 5593879

; GENERAL INFORMATION:

; APPLICANT: Steller, Hermann

; APPLICANT: Abrams, John M.

; APPLICANT: Grether, Megan E.

; APPLICANT: White, Kristin

; TITLE OF INVENTION: Cell Death Genes of Drosophila

; TITLE OF INVENTION: Melanogaster and Vertebrate Analogs

; NUMBER OF SEQUENCES: 16

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.

; STREET: Two Militia Drive

; CITY: Lexington

; STATE: MA

; COUNTRY: US

; ZIP: 02173

; COMPUTER READABLE FORM:

; MEDIUM TYPE: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/123,343A

; FILING DATE: 17-SEP-1993

; CLASSIFICATION: 800

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/004,957

; FILING DATE: 15-JAN-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: Granahan, Patricia

; REGISTRATION NUMBER: 32,227

; REFERENCE/DOCKET NUMBER: MIT-5907A

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 617-861-6240

; TELEFAX: 617-861-9540

; INFORMATION FOR SEQ ID NO: 15:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 503 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: unknown

; TOPOLOGY: unknown

; MOLECULE TYPE: DNA (genomic)

; US-08-123-343A-15

Db 3709 CTCAACGGCACTGTTCCGGGGGGCTTTTCAATCTACCGTTAGTTACGATTATCGAATC 3768
QY 1196 accgggtgaacattaccgggatgtagatcggggtaagtgttaagtccaacaacatcgctatt 1255
Db 3769 ACTGATAATTTCTTCTCCGGTGAACCTCCGGTAACGATGTCGGCGGATGTTCTCGATCAG 3828
QY 1256 cttaacagctaaagcagggtgcataatggggccagatacacaccatcaat 1302
Db 3829 ATTTACCTCTCTAACAACAGTGGTTTTCGGCGGAGATTCCACCTCGCAT 3875

RESULT 4
US-08-633-770A-3/c
; Sequence 3, Application US/08633770A
; Patent No. 5908760
; GENERAL INFORMATION:
; APPLICANT: Bojsen, Kirsten
; APPLICANT: Yu, Shukun
; APPLICANT: Kragh, Karsten
; APPLICANT: Christensen, Tove
; APPLICANT: Marcussen, Jan
; TITLE OF INVENTION: ALPHA-1,4-GLUCAN LYASE FROM A FUNGUS, ITS
; TITLE OF INVENTION: PURIFICATION GENE CLONING AND EXPRESSION IN MICROORGANISMS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/633,770A
; FILING DATE: July 8, 1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP94/03398
; FILING DATE: OCT-15-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Altman, Daniel E
; REGISTRATION NUMBER: 34,115
; REFERENCE/DOCKET NUMBER: DY06.001APC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 714-760-0404
; TELEFAX: 714-760-9502
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3201 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-633-770A-3

Query Match 2.3%; Score 33.2; DB 3; Length 3201;
Best Local Similarity 50.0%; Pred. No. 0.84;
Matches 83; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

QY 243 ccggggcgtaaacgggtaacgcatacactgcagcgaagaatacgaaaaagagtttcattaca 302
Db 1028 CTGGGCTTCAACTTTGAACGCTCCAACAGACCTGTATAAAGTCGTACAAATTCGGGACC 969
QY 303 ccattggcccttattccacaccggtcctgaccatcgaaacccgggtgacggattatttcg 362
Db 968 GTATCCGCACTGATACCGTAACCAATCGATACCAACCATACCTGTTCCCGACGTTGATG 909

QY 363 acactcgagatgcttttgaagggtgctatcaattcgggaacaggatat 408
Db 908 CCTGAGTTGCTCTTCCAAAGTCGATGGCAATTTGAGAGTAGTTAT 863

RESULT 5
US-08-633-770A-11/c
; Sequence 11, Application US/08633770A
; Patent No. 5908760
; GENERAL INFORMATION:
; APPLICANT: Bojsen, Kirsten
; APPLICANT: Yu, Shukun
; APPLICANT: Kragh, Karsten
; APPLICANT: Christensen, Tove
; APPLICANT: Marcussen, Jan
; TITLE OF INVENTION: ALPHA-1,4-GLUCAN LYASE FROM A FUNGUS, ITS
; TITLE OF INVENTION: PURIFICATION GENE CLONING AND EXPRESSION IN MICROORGANISMS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/633,770A
; FILING DATE: July 8, 1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP94/03398
; FILING DATE: OCT-15-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Altman, Daniel E
; REGISTRATION NUMBER: 34,115
; REFERENCE/DOCKET NUMBER: DY06.001APC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 714-760-0404
; TELEFAX: 714-760-9502
; TELEX:
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4726 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-633-770A-11

Query Match 2.3%; Score 33.2; DB 3; Length 4726;
Best Local Similarity 50.0%; Pred. No. 1.1;
Matches 83; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

QY 243 ccggggcgtaaacgggtaacgcatacactgcagcgaagaatacgaaaaagagtttcattaca 302
Db 1741 CTGGGCTTCAACTTTGAACGCTCCAACAGACCTGTATAAAGTCGTACAAATTCGGGACC 1682
QY 303 ccattggcccttattccacaccggtcctgaccatcgaaacccgggtgacggattatttcg 362
Db 1681 GTATCCGCACTGATACCGTAACCAATCGATACCAACCATACCTGTTCCCGACGTTGATG 1622
QY 363 acactcgagatgcttttgaagggtgctatcaattcgggaacaggatat 408
Db 1621 CCTGAGTTGCTCTTCCAAAGTCGATGGCAATTTGACAGTAGTTAT 1576

RESULT 6


```
Db 508 GATGGAGCTATATTGA 492
|||||
RESULT 10
US-08-374-686-1/c
; Sequence 1, Application US/08374686
; Patent No. 5616474
; GENERAL INFORMATION:
; APPLICANT: Bolotin, Monique
; APPLICANT: Menart, Sandrine
; TITLE OF INVENTION: K. lactis Transaldolase Gene Promoter
; TITLE OF INVENTION: and Use Thereof
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rhone-Poulenc Rorer Inc.
; STREET: 500 Arcola Rd. 3C43
; CITY: Collegeville
; STATE: PA
; COUNTRY: USA
; ZIP: 19002
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/374,686
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/FR93/00771
; FILING DATE: 28-JUL-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 92/09432
; FILING DATE: 30-JUL-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, Julie K.
; REGISTRATION NUMBER: 38,619
; REFERENCE/DOCKET NUMBER: ST92048-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (610)454-3839
; TELEFAX: (610)454-3808
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1349 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Kluyveromyces lactis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1297..1347
US-08-374-686-1

Query Match 2.2%; Score 32.2; DB 1; Length 1349;
Best Local Similarity 63.6%; Pred. No. 1.1;
Matches 49; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 1160 ctgaacaaaacctgttagtttagtagaataactaacccggtgaacattaccggatgta 1219
|||||
Db 640 GCGATGACACACCGGGTAGTTTATGCGGGATGACTTTCCAGTGTATCCTTGCCTGGTGGTA 581
|||||
QY 1220 gatcggggtaatgtga 1236
|||||
Db 580 GATCGAGCTATATTGA 564
|||||

RESULT 11
US-08-743-637B-16
; Sequence 16, Application US/08743637B
; Patent No. 5994066
; GENERAL INFORMATION:
; APPLICANT: BERGERON, Michel G.
; APPLICANT: PICARD, Francois J.
; APPLICANT: OUELLETTE, Marc
; APPLICANT: ROY, Paul H.
; TITLE OF INVENTION: SPECIES-SPECIFIC AND UNIVERSAL DNA
; TITLE OF INVENTION: PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND
; TITLE OF INVENTION: IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED
; TITLE OF INVENTION: ANTIBIOTIC RESISTANCE GENES FROM CLINICAL SPECIMENS
; NUMBER OF SEQUENCES: 273
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: QUARLES & BRADY
; STREET: 411 EAST WISCONSIN AVENUE
; CITY: MILWAUKEE
; STATE: WISCONSIN
; COUNTRY: USA
; ZIP: 53202-4497
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/743,637B
; FILING DATE: 04-NOV-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/526,840
; FILING DATE: 11-SEP-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: BAKER, Jean C.
; REGISTRATION NUMBER: 35,433
; REFERENCE/DOCKET NUMBER: 850586.90012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (414) 277-5000
; TELEFAX: (414) 277-5591
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2167 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Pseudomonas aeruginosa
US-08-743-637B-16

Query Match 2.2%; Score 31.2; DB 4; Length 2167;
Best Local Similarity 51.4%; Pred. No. 3;
Matches 72; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 1303 attgggttactttactcttcacgagtgacggcgacacaaagtgtgcacaaatggcgc 1362
|||||
Db 162 ATCGGCTGCTGTCTCCGAAACCGCGTCACCGCGGATATCGAGCGCTCGCACGCGTAT 221
|||||
QY 1363 ggagcaaccaggctattgccgaaataatcaaaatggcggaatcaacgcagaccactc 1422
|||||
Db 222 GCGCATTCGTCGCGGTGAGCAACTGAACCGCGAGGCGGCGTCCGCCGATC 281
|||||
QY 1423 aatgcattcatttgatcc 1442
|||||
Db 282 GAAACGCTGCCAGGACCC 301
|||||

RESULT 12
US-08-526-840B-16
; Sequence 16, Application US/08526840B
; Patent No. 6001564
; GENERAL INFORMATION:
```

```
; APPLICANT: BERGERON, Michel G.
; APPLICANT: OUELLETTE, Marc
; APPLICANT: ROY, Paul H.
; TITLE OF INVENTION: SPECIFIC AND UNIVERSAL PROBES AND
; TITLE OF INVENTION: AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND IDENTIFY
; TITLE OF INVENTION: COMMON BACTERIAL PATHOGENS AND ANTIBIOTIC RESISTANCE GENES
; TITLE OF INVENTION: FROM CLINICAL SPECIMENS FOR ROUTINE DIAGNOSIS IN ...
; NUMBER OF SEQUENCES: 177
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: QUARLES & BRADY
; STREET: 411 East Wisconsin Avenue
; CITY: Milwaukee
; STATE: Wisconsin
; COUNTRY: USA
; ZIP: 53202-4497
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/526,840B
; FILING DATE: 11-SEP-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/304,732
; FILING DATE: 12-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BAKER, Jean C.
; REGISTRATION NUMBER: 35,433
; REFERENCE/DOCKET NUMBER: 850586.90012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (414) 277-5000
; TELEFAX: (414) 277-5591
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2167 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Pseudomonas aeruginosa
; US-08-526-840B-16

Query Match 2.2%; Score 31.2; DB 5; Length 2167;
Best Local Similarity 51.4%; Pred. No. 3;
Matches 72; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 1303 attggtttacttactccttcaggagtgacggcgacaaagattgtcacaaatggcgc 1362
Db 162 ATCGGCTGCTGTCTCCGAACCGCGTCACCGCGATATCGAGCGCTCGCAGCGTAT 221
QY 1363 gagacaccacagcatttgcgcgaataatcaaaatggcgcatcaacggcagaccactc 1422
Db 222 GCGCATTCGTCGCGTGAGCAACTGAACCGCGAGCGCGCTCGCGCGTCCCGCATC 281
QY 1423 aatgcaattcatttgatcc 1442
Db 282 GAAACGCTGTCCAGGACCC 301

RESULT 13
; Sequence 2, Application PC/TUS9513536
; GENERAL INFORMATION:
; APPLICANT: SKOPEK, THOMAS R
; TITLE OF INVENTION: SYNTHESIS OF METHYLASE RESISTANT GENES
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ELLIOTT KORSEN
; STREET: 126 E. Lincoln Avenue, P.O. Box 2000
```

```
; CITY: Rahway
; STATE: NJ
; COUNTRY: US
; ZIP: 07065
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/13536
; FILING DATE: 13-OCT-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: KORSEN, ELLIOTT
; REGISTRATION NUMBER: 32,705
; REFERENCE/DOCKET NUMBER: 19257Y
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908) 594-5493
; TELEFAX: (908) 594-4720
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1187 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: Internal
; PCT-US95-13536-2

Query Match 2.1%; Score 30.8; DB 6; Length 1187;
Best Local Similarity 55.7%; Pred. No. 2.9;
Matches 59; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 822 ctggagggccctgtttattggatgcccattgttcagggtgagtgatgagattgcg 881
Db 868 CTGAATCTGGCTGCGTGTGGTGACAGATATCTGTGGTGGGCTATGATGATGACTGAAG 927
QY 882 ggaccgcagtagatttgccctcaatcacaccaccatcaaaagtcgattt 927
Db 928 ACAGCTCATGTTATATCCACCAAGACCAACCAACAGGATTT 973

RESULT 14
; Sequence 1, Application PC/TUS9513536
; GENERAL INFORMATION:
; APPLICANT: SKOPEK, THOMAS R
; TITLE OF INVENTION: SYNTHESIS OF METHYLASE RESISTANT GENES
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ELLIOTT KORSEN
; STREET: 126 E. Lincoln Avenue, P.O. Box 2000
; CITY: Rahway
; STATE: NJ
; COUNTRY: US
; ZIP: 07065
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/13536
; FILING DATE: 13-OCT-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: KORSEN, ELLIOTT
; REGISTRATION NUMBER: 32,705
; REFERENCE/DOCKET NUMBER: 19257Y
```

TELECOMMUNICATION INFORMATION:
TELEPHONE: (908) 594-5493
TELEFAX: (908) 594-4720
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1188 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: internal
PCR-US95-13536-1

Query Match 2.1%; Score 30.8; DB 6; Length 1188;
Best Local Similarity 55.7%; Pred. No. 2.9;
Matches 59; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 822 ctggaggccgcctgtttattgtgtatgccccatcttctcagggtgatgtgagatttgcg 881
DB 868 CTGAATCTGGCCTGCGTGTGGTGCAGATATCTCTGTGTGGGCTATGATGACTGAAG 927
QY 882 ggaccgcagtagatttgctcaatcacaccatcaaaagtcgattt 927
DB 928 ACAGCTCATGTTATATCCCAACCAAGCACCACCATCAACAGGATTT 973

RESULT 15

US-08-822-586-45
Sequence 45, Application US/08822586
Patent No. 6015890

GENERAL INFORMATION:
APPLICANT: WILLIAM R. JACOBS, JR., JAMES M. MUSSER AND
APPLICANT: AMALIO TELENTE
TITLE OF INVENTION: AN EMBAC OPERON OF MYCOBACTERIA AND
TITLE OF INVENTION: MUTANTS THEREOF
NUMBER OF SEQUENCES: 57
CORRESPONDENCE ADDRESS:
ADDRESSER: AMSTER, ROTHSTEIN & EBENSTEIN
STREET: 90 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: U.S.A.
ZIP: 10016

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH 1.44 Mb STORAGE
MEDIUM TYPE: DISKETTE
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/822,586
FILING DATE: MARCH 20, 1997

ATTORNEY/AGENT INFORMATION:

NAME: ELIZABETH A. BOGOSIAN
REGISTRATION NUMBER: 39,911
REFERENCE/DOCKET NUMBER: 96700/437
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 697-5995

TELEFAX: (212) 286-0854 or 286-0082
TELEX: TWX 710-581-4766

INFORMATION FOR SEQ ID NO: 45:
SEQUENCE CHARACTERISTICS:

LENGTH: 10095
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
HYPOTHETICAL: NO

US-08-822-586-45

Query Match 2.1%; Score 30.2; DB 5; Length 10095;
Best Local Similarity 49.7%; Pred. No. 17;
Matches 77; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 901 ctcaatcacccaccatcaaaagtctgatttgatcaagaactggcagcttctcctggccacgaat 960
DB 2958 CCCGAGCACCCGCCACCAGTCCGGTTGTCGCCGACGACAGGATCTGGCCCGCAGCA 3017
QY 961 ggagaatgccgaaaaatattattgagttattggcagtgccactccgctggaggtatgcgacgcg 1020
DB 3018 CTGGATCGCCCTCACACCACCGCGGATTCGGGGGTGCGCAGCTGCAGAACGTGTGTGG 3077
QY 1021 aattgcataatcgacctaatttacttgctgctgtag 1055
DB 3078 CGCAGCGGATCCGGTGTCTCTGGACTGGCTGGTGG 3112

Search completed: September 9, 2000, 20:47:22
Job time: 1875 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 9, 2000, 20:02:07 ; Search time 588.37 Seconds
(without alignments)
10807.778 Million cell updates/sec

Title: US-09-214-679-1
Perfect score: 1442
Sequence: 1 ccgggaatccatgtgcc.....aatgcaattcattggatcc 1442

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5247842 seqs, 2204914090 residues
Total number of hits satisfying chosen parameters: 10495684

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*

44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_est45:*
80: gb_est46:*
81: gb_est47:*
82: em_est35:*
83: em_est36:*
84: em_est37:*
85: gb_est48:*
86: gb_est49:*
87: gb_est50:*
88: gb_est51:*
89: gb_est52:*
90: gb_est53:*
91: gb_est54:*
92: gb_est55:*
93: gb_gss1:*
94: gb_gss2:*
95: gb_gss3:*
96: gb_gss4:*
97: em_gss1:*
98: em_gss2:*
99: em_gss3:*
100: em_gss4:*
101: gb_gss5:*
102: gb_gss6:*
103: gb_gss7:*
104: gb_gss8:*
105: gb_gss9:*
106: em_gss5:*
107: em_gss6:*
108: em_gss7:*
109: em_gss8:*
110: em_gss9:*
111: em_gss10:*
112: em_gss11:*
113: gb_gss10:*
114: gb_gss11:*
115: em_gss12:*
116: gb_gss12:*

IMAGE:232927 5', mRNA sequence.

ACCESSION H75593.1 GI:1049521
 VERSION H75593.1
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 421)
 AUTHORS Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B., Chissoe, S., Dietrich, N., DuBuque, T., Favello, A., Gish, W., Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevaskis, E., Underwood, K., Wohldmann, P., Waterston, R., Wilson, R. and Marra, M.

TITLE Generation and analysis of 280,000 human expressed sequence tags
 JOURNAL Genome Res. 6 (9), 807-828 (1996)
 MEDLINE 97044478
 COMMENT On Mar 12, 1999 this sequence version replaced gi:4057371.
 Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 Insert Size: 1012
 High quality sequence stops: 366
 Source: IMAGE Consortium, LLNL
 This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 1012 Std Error: 0.00
 Seq primer: M13RP1
 High quality sequence stop: 366.
 Location/Qualifiers
 1..421
 /organism="Homo sapiens"
 /db_xref="GDB:3786104"
 /db_xref="taxon:9606"
 /clone="IMAGE:232927"
 /clone_lib="Soares fetal liver spleen INFLS"
 /sex="male"
 /dev_stage="20 week-post conception fetus"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site.1: Pac I; Site.2: Eco RI; 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5', AACTGGAGATTAATAAGATCTTTTATTTTATTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 89 a 104 c 107 g 113 t 8 others
 ORIGIN

Query Match 2.4%; Score 35; DB 86; Length 421;
 Best Local Similarity 51.6%; Pred. No. 4.1;
 Matches 80; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 1044 actggctgtagaagactttggcttcgaacaatgggagctacatgcttctgagtcatt 1103
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 242 ACTGGGAAGAGATGATCTTTGGCTCGGAGGAGATGATCTCAACGGAGGCTGAGACAT 183

QY 1104 gcggcaaatgcgctgggcaacatggtgcaccccaataacacacctggtgcgcatgctga 1163
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 182 CCCACAGGAGGATCTTGGGGCGGAATGCGCGTGCACAGTTACACCGAGTTGGGGTGAACC 123

QY 1164 acaaaaacctgttagtttagtaagaataaactaac 1198
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 122 ACTATACCTTTATATAGAGTGTATCTTTCTATCCC 88

QY	535	ctacggcaatcgtgcacatgattccgatcttatttgcyggaactaccggagccgacctgacggc	594
DG	167	CAGAGACTTCTTCAGCATGAAGCGGGAGTGGGAGAACCTTGCAACCAGTCCAACGTGCGGCG	108
QY	595	catcgtaaatgatccgctgccagaaaaggigtcgcatgataaaactcgacagtgaagaagt	654
DG	107	CATGCACACTNGCTGCGGCTGAACGAGGTGATCCTGAAGAAATCCCGGGACGCCAAGCT	48
QY	655	ctactggagcaaacgcgcatacgcgtcttcccataaaacccccattg	698
DG	47	TGTTTTCCTCAACATGCTGGCGCTCCCGCACCGCAATGTG	4
RESULT	7		
B57961			
LOCUS	B57961	691 bp DNA	GSS 20-JUN-1998
DEFINITION	CIT-HSP-2011C6.TRB CIT-HSP Homo sapiens genomic clone 2011C6, genomic survey sequence.		
ACCESSION	B57961		
VERSION	B57961.1	GI:2612295	
KEYWORDS	GSS.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and Venter,J.C. Use of a random BAC End Sequence Database for Sequence-Ready Map Building Unpublished (1997) On Dec 15, 1999 this sequence version replaced gi:4575880. Contact: Mark Adams Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: mdadams@tigr.org Clones are available from Research Genetics (info@resgen.com). BAC end search page: http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html Seq primer: M13 Reverse Class: BAC ends.		
TITLE	Location/Qualifiers		
JOURNAL	1. 691		
COMMENT	/organism="Homo sapiens" /db_xref="GDB:7041464" /db_xref="taxon:9606" /clone="2011C6" /clone_lib="CIT-HSP" /sex="Male" /cell_type="Sperm" /note="Vector: pBelOBAC11; Site_1: HindIII; Site_2: HindIII"		
FEATURES	source		
	227 a 86 c 153 g 225 t		
BASE COUNT			
ORIGIN			
Query Match	2.4%; Score 35;	DB 120;	Length 691;
Best Local Similarity	52.4%;	Pred. No. 4.9;	
Matches	77; Conservative	0; Mismatches	70; Indels 0; Gaps 0;
QY	1122	gcaacatggtcgaccccaatacacaccgttgccgcatgtagtaacaaaaacacctgttagtt	1181
DG	423	GCAGTGTGGTCTACTAGAAAATAGAATGTGATCGCCCTGTGTAATTTTAAATTTTAAATTT	482
QY	1182	agtaggaataactaacccggtgaacattaccocggatgtagatcgcgggtaatgctgaagtgc	1241
DG	483	TC TAGTAAACCACTAATAAGACAGAAACATTAAATTTTAAATAGTGTATTTATTTAATCC	542

QY 1242 aaacaatcgctatttttaacagctaaa 1268
 Db 543 AATATATCCAAATAATGATCATTTCAA 569

RESULT 8
 A0610392/c
 LOCUS
 DEFINITION HS_5101_B2_A01_SP6E RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate=677 Col=2 Row=B, genomic survey sequence.
 ACCESSION A0610392

VERSION A0610392.1 GI:5071668
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 396)
 AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Kelller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
 TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
 JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 MEDLINE 99380589
 COMMENT On Feb 19, 1999 this sequence version replaced gi:4146198.
 Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Resear h Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu
 Plate: 677 row: B column: 2
 Seq primer: SP6
 Class: BAC ends
 High quality sequence stop: 396.
 Location/Qualifiers
 1..396
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="Plate=677 Col=2 Row=B"
 /clone_lib="RPCI-11 Human Male BAC Library"
 /sex="male"
 /note="Vector: pBAC3.6; Site.1: EcoRI; Site.2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBAC3.6 vector at EcoRI sites"
 BASE COUNT 128 a 113 c 66 g 89 t
 ORIGIN

Query Match 2.4%; Score 34.8; DB 105; Length 396;
 Best Local Similarity 47.3%; Pred. No. 4.7;
 Matches 105; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

QY 827 ggccgctgtttattggtgcccattgctgtcagggtgatgtgagattgcgggacc 886
 Db 358 GGCCGGGGTTGTGTGTGTGTCCTCCATGTCATTTGGTTAGGCCAGCTTGGGGTTT 299

QY 887 gcagtagattgctcaatcacaccatcaagaagtcgattgatcaagaactggcagctt 946
 Db 298 ACAGAGGCGCTGGCCACACAGGTCGGAAGGAATGGTGTGAGTTTCTTTGGGGTTTAGGGTT 239

QY 947 tcttgccacgaatggagaatgccgaaataattattagttgagtcgacgtccgctg 1006
 Db 238 TGGGCTCTACATTAGTGAACCTTTGGATAATTTTATAGGGTAGAGGCTCGGTCCGCTT 179

QY 1007 gaggatgcgacgcgaattgcataatcgacttaattactgg 1048
 Db 178 TACCAAGGTAGCAGTTCACAGTGCATTTATTAATAGG 137

RESULT 9
 A0555596/c
 LOCUS
 DEFINITION HS_5229_A1_A09_SP6E RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate=805 Col=17 Row=A, genomic survey sequence.
 ACCESSION A0555596

VERSION A0555596.1 GI:4915603
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 422)
 AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Kelller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
 TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
 JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 MEDLINE 99380589
 COMMENT On Dec 15, 1999 this sequence version replaced gi:4215100.
 Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Resear h Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu
 Plate: 805 row: A column: 17
 Seq primer: SP6
 Class: BAC ends
 High quality sequence stop: 422.
 Location/Qualifiers
 1..422
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="Plate=805 Col=17 Row=A"
 /clone_lib="RPCI-11 Human Male BAC Library"
 /sex="male"
 /note="Vector: pBAC3.6; Site.1: EcoRI; Site.2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBAC3.6 vector at EcoRI sites"
 BASE COUNT 143 a 76 c 58 g 144 t
 ORIGIN

Query Match 2.4%; Score 34.8; DB 104; Length 422;
 Best Local Similarity 49.5%; Pred. No. 4.8;
 Matches 90; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 587 ctgacggccatgctcaatgatccgtgcagaaagtgcgcatgataaactcgacgt 646
 Db 421 CAGTAGGTAGCGCTAATGCTCAATATCCTTAATTGTAGGTAGTATAAACTAAAGTGT 362

Qy	567	gcgagctgacgggacgcgacctgacggccatgct 600
Db	261	GCACACTGACCGTGAAGGCCCCCATGCCAAGCT 228
RESULT 11		
LOCUS	AI346857/c	
DEFINITION	AI346857 477 bp mRNA EST 30-DEC-1998	
	qp53612.x1 NCI_CGAP_Co8 Homo sapiens cDNA clone IMAGE:1926766 3',	
	similar to gb:223090 HEAT SHOCK 27 KD PROTEIN (HUMAN);, mRNA	
	sequence.	
ACCESSION	AI346857	
VERSION	AI346857.1	GI:4084063
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
TITLE	1. (Bases 1 to 477)	
	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .	
	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),	
	Tumor Gene Index	
JOURNAL	Unpublished (1997)	
COMMENT	On Apr 7, 1998 this sequence version replaced gi:3036231.	
	Contact: Robert Strausberg, Ph.D.	
	Tel: (301) 496-1550	
	Email: Robert_Strausberg@nih.gov	
	Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.	
	Emmert-Buck, M.D., Ph.D.	
	CDNA Library Preparation: M. Bento Soares, Ph.D.	
	CDNA Library Arrayed by: Greg Lennon, Ph.D.	
	DNA Sequencing by: Washington University Genome Sequencing Center	
	clone distribution: NCI-CGAP clone distribution information can be	
	found through the I.M.A.G.E. Consortium/LLNL at:	
	www-bio.llnl.gov/bbrp/image/image.html	
Trace considered overall poor quality		
	Seq primer: -400P from Gibco	
	High quality sequence stop: 1.	
FEATURES	Location/Qualifiers	
source	1..477	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/clone="IMAGE:1926766"	
	/clone_lib="NCI_CGAP_Co8"	
	/tissue_type="adenocarcinoma"	
	/lab_host="DH10B"	
	/note="Organ; Vector: pT7T3D-Pac (Pharmacia) with a	
	modified polylinker; 1st strand cDNA was prepared from	
	colon adenocarcinoma, and was then primed with a Not I -	
	oligo(dT) primer. Double-stranded cDNA was ligated to Eco	
	RI adaptors (Pharmacia), digested with Not I and cloned	
	into the Not I and Eco RI sites of the modified pT7T3	
	vector. Library is normalized. Library was constructed by	
	Bento Soares and M. Fatima Bonaldo.	
BASE COUNT	93 a 118 c 152 g 114 t	
ORIGIN		
Query Match 2.4% Score 34.8; DB 38; Length 477;		
Best Local Similarity 60.6%; Pred. No. 5;		
Matches 57; Conservative 0; Mismatches 37; Indels 0; Gaps 0;		
Qy	507	aatccatgttgcccgcgcggttatccatccgacgcatctgcgccatgattccgcatttg 566
Db	317	AATACACGCTGCCCGCGGTGGACACCCACCAAGTTTCTCTCTCCCTGTGCCATGAGG 258
Qy	567	gcgagctgacgggacgcgacctgacggccatgct 600
Db	257	GCACACTGACCGTGAAGGCCCCCATGCCAAGCT 224

[W][O][R][D] [S][E][Q][U][E][N][C][E] [A][N][A][L][Y][S][I][S]

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Thu Aug 24 18:54:30 2000; MasPar time 10.13 Seconds
Tabular output not generated. 767.194 Million cell updates/sec

Title: >US-09-214-679-2
Description: (1-328) from US09214679.pep
Perfect Score: 2417
Sequence: 1 MKWLEESIMAKRGVAGRK...GNMVDPKYTVGAMLNKLLV 328

Scoring table: PAM 150
Gap 11

Searched: 188963 seqs, 23686106 residues
Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq36
1:geneseqp

Statistics: Mean 34.490; Variance 153.853; scale 0.224

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match Length DB ID	Description Pred. No.
1	2393	99.0	328 1 W40263 K. oxytoca R-specific 1.62e-213
2	200	8.3	33 1 W40275 K. oxytoca R-specific 3.33e-07
3	157	6.5	21 1 W40272 K. oxytoca R-specific 1.05e-03
4	148	6.1	20 1 W40264 K. oxytoca R-specific 5.37e-03
5	132	5.5	18 1 W40270 K. oxytoca R-specific 9.25e-02
6	132	5.5	575 1 W58856 C. acidivorans gamma-1 9.25e-02
7	122	5.0	19 1 W40271 K. oxytoca R-specific 5.23e-01
8	122	5.0	23 1 W40273 K. oxytoca R-specific 5.23e-01
9	107	4.4	482 1 W08199 B.licheniformis mutant 6.51e+00
10	107	4.4	561 1 W97413 Lucilia cuprina GABA r 6.51e+00
11	104	4.3	14 1 W40274 K. oxytoca R-specific 1.06e+01
12	102	4.2	483 1 R24250 Alpha amylase. 1.47e+01
13	102	4.2	483 1 R34289 B.licheniformis alpha 1.47e+01
14	102	4.2	512 1 W10325 B.licheniformis acid 2.03e+01
15	100	4.1	483 1 W08206 B.licheniformis mutant 2.03e+01
16	100	4.1	483 1 W05230 Bacillus licheniformis 2.03e+01
17	100	4.1	483 1 W58737 M15E alpha-amylase. 2.03e+01
18	100	4.1	483 1 R58721 M197E alpha-amylase. 2.03e+01
19	100	4.1	483 1 R58003 M366A alpha-amylase. 2.03e+01
20	100	4.1	483 1 W08193 B.licheniformis mutant 2.03e+01
21	100	4.1	483 1 W39742 B.licheniformis alpha 2.03e+01
22	100	4.1	483 1 W08211 B.licheniformis mutant 2.03e+01
23	100	4.1	483 1 W31404 Bacillus licheniformis 2.03e+01

ALIGNMENTS

RESULT 1			
ID	W40263	standard; Protein; 328 AA.	
AC	W40263		
DT	16-JUN-1998	(first entry)	
DE	K. oxytoca R-specific amidohydrolase protein.		
KW	R-specific amidohydrolase; hydrolysis; racemate; nitrogen source; 3,3,3-trifluoro-2-hydroxy-2-methyl propionamide; isoform.		
OS	Klebsiella oxytoca.		
PN	W09801568-A2.		
PD	15-JAN-1998.		
PF	10-JUL-1997; E03670.		
PR	03-MAR-1997; CH-000500.		
PR	10-JUL-1996; CH-001723.		
PA	(LONZ) LONZA AG.		
PI	Brieden W, Naughton A, Robins K, Shaw N, Tinschert A, Zimmermann T;		
PI	WPI; 98-101063/09.		
DR	N-PSDB; V10449.		
PT	Tri:fluoro-2-hydroxy-2-methyl-propionic acid enantiomer preparation		
PT	- by stereoselective hydrolysis of corresponding racemic amide using		
PT	microorganism or derived enzyme, used as drug intermediate		
PS	Claim 5; Page 34-35; 68pp; German.		
CC	This sequence represents a R-specific amidohydrolase isolated from		
CC	Klebsiella oxytoca strain PRS1 which allows the microorganism to utilise		
CC	3,3,3-trifluoro-2-hydroxy-2-methyl propionamide as its sole nitrogen		
CC	source. This amidohydrolase is used in a process for preparing		
CC	(R)-isoforms of 3,3,3-trifluoro-2-hydroxy-2-methyl propionamide which		
CC	is cheaper than prior art optical resolution of the racemate using		
CC	dimethoxy strychnine or (S)-(-)-alpha-methylbenzylamine.		
SC	Sequence 328 AA;		

Query Match 99.08; Score 2393; DB 1; Length 328;			
Best Local Similarity 99.4%; Pred. No. 1.62e-213;			
Matches 326; Conservative 0; Mismatches 2; Indels 0; Gaps 0;			
Db	1	MKWLEESIMAKRGVAGRKPVTHLTEMQKEHYTIGPSTPVLTEPDRIVDTRDA	60
Qy	1	MKWLEESIMAKRGVAGRKPVTHLTEMQKEHYTIGPSTPVLTEPDRIVDTRDA	60
Db	61	FEAGAINSEQDIPQLKMPFNQNGIMVNGAEKGDVLAIVIESMLPRGVPGICAMI	120
Qy	61	FEAGAINSEQDIPQLKMPFNQNGIMVNGAEKGDVLAIVIESMLPRGVPGICAMI	120
Db	121	PHFGGLTGDTLMTLNDPLPEKVRMIKLDSEKVKYWSRHTLPYKPHGTGLSVSPEDSIN	180
Qy	121	PHFGGLTGDTLMTLNDPLPEKVRMIKLDSEKVKYWSRHTLPYKPHGTGLSVSPEDSIN	180

Db 181 SLTPDHGNDVDPDIPGSLTYLPLVRAPGRLFIGDAHACQGGTICGTAVEFASITTI 240
 QY 181 SLTPDHGNDVDPDIPGSLTYLPLVRAPGRLFIGDAHACQGGTICGTAVEFASITTI 240
 Db 241 KVDLKNWOLSWPRMENAEMISGSRPLEDATRIAYRDLIYWLVEDFGEQWDAYMLL 300
 QY 241 KVDLKNWOLSWPRMENAEMISGSRPLEDATRIAYRDLIYWLVEDFGEQWDAYMLL 300
 Db 301 SQCGKVRGNNVDPKYTVGMLNKNLLV 328
 QY 301 SQCGKVRGNNVDPKYTVGMLNKNLLV 328

RESULT 2
 ID W40275 standard; Protein; 33 AA.
 AC W40275; 16-JUN-1998 (first entry)
 DT 16-JUN-1998 (first entry)
 DE K. oxytoca R-specific amidohydrolase peptide T13-3.
 KW R-specific amidohydrolase; hydrolysis; racemate; nitrogen source;
 OS 3,3,3-trifluoro-2-hydroxy-2-methyl propionamide; isoform.
 PN Klebsiella oxytoca.
 PD WO9801568-A2.
 PF 15-JAN-1998.
 PR 10-JUL-1997; E03670.
 PR 03-MAR-1997; CH-000500.
 PR 10-JUL-1996; CH-001723.
 PA (LONZ) LONZA AG.
 PI Brieden W, Naughton A, Robins K, Shaw N, Tinschert A,
 PI Zimmermann T;
 DR WPI; 98-101063/09.
 DR N-PSDB; V10452.
 PT Tri:fluoro-2-hydroxy-2-methyl-propionic acid enantiomer preparation
 PT - by stereoselective hydrolysis of corresponding racemic amide using
 PT microorganism or derived enzyme, used as drug intermediate
 PS Example 10.2; Page 29; 68pp; German.
 CC W40264-W40275 are peptide fragments of an R-specific amidohydrolase
 CC isolated from Klebsiella oxytoca strain PRS1. This enzyme allows the
 CC microorganism to utilize 3,3,3-trifluoro-2-hydroxy-2-methyl propionamide
 CC as its sole nitrogen source. This amidohydrolase is used in a process for
 CC preparing (R)-isoforms of 3,3,3-trifluoro-2-hydroxy-2-methyl propionamide
 CC which is cheaper than prior art optical resolution of the racemate using
 CC dimethoxy strychnine or (S)-(-)-alpha-methylbenzylamine.
 SQ Sequence 33 AA;

Query Match 8.3%; Score 200; DB 1; Length 33;
 Best Local Similarity 87.9%; Pred. No. 3.33e-07; Indels 0; Gaps 0;
 Matches 29; Conservative 0; Mismatches 4;

Db 1 GVDYPGIEAMIPHFGLGTDLTAMLDLPK 33
 QY 110 GVDYPGIEAMIPHFGLGTDLTAMLDLPK 142

RESULT 3
 ID W40272 standard; Protein; 21 AA.
 AC W40272; 16-JUN-1998 (first entry)
 DT 16-JUN-1998 (first entry)
 DE K. oxytoca R-specific amidohydrolase peptide T9-2.
 KW R-specific amidohydrolase; hydrolysis; racemate; nitrogen source;
 OS 3,3,3-trifluoro-2-hydroxy-2-methyl propionamide; isoform.
 PN Klebsiella oxytoca.
 PD WO9801568-A2.
 PF 15-JAN-1998.
 PR 10-JUL-1997; E03670.
 PR 03-MAR-1997; CH-000500.
 PR 10-JUL-1996; CH-001723.
 PA (LONZ) LONZA AG.
 PI Brieden W, Naughton A, Robins K, Shaw N, Tinschert A,
 PI Zimmermann T;
 DR WPI; 98-101063/09.
 PT Tri:fluoro-2-hydroxy-2-methyl-propionic acid enantiomer preparation
 PT - by stereoselective hydrolysis of corresponding racemic amide using
 PT microorganism or derived enzyme, used as drug intermediate

PS Example 10.2; Page 29; 68pp; German.
 CC W40264-W40275 are peptide fragments of an R-specific amidohydrolase
 CC isolated from Klebsiella oxytoca strain PRS1. This enzyme allows the
 CC microorganism to utilize 3,3,3-trifluoro-2-hydroxy-2-methyl propionamide
 CC as its sole nitrogen source. This amidohydrolase is used in a process for
 CC preparing (R)-isoforms of 3,3,3-trifluoro-2-hydroxy-2-methyl propionamide
 CC which is cheaper than prior art optical resolution of the racemate using
 CC dimethoxy strychnine or (S)-(-)-alpha-methylbenzylamine.
 SQ Sequence 21 AA;

Query Match 6.5%; Score 157; DB 1; Length 21;
 Best Local Similarity 100.0%; Pred. No. 1.05e-03;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 EFHYTIGPYSTPVLTIPEGDR 21
 QY 32 EFHYTIGPYSTPVLTIPEGDR 52

RESULT 4
 ID W40264 standard; Protein; 20 AA.
 AC W40264; 16-JUN-1998 (first entry)
 DT 16-JUN-1998 (first entry)
 DE K. oxytoca R-specific amidohydrolase N-terminal peptide.
 KW R-specific amidohydrolase; hydrolysis; racemate; nitrogen source;
 OS 3,3,3-trifluoro-2-hydroxy-2-methyl propionamide; isoform.
 PN Klebsiella oxytoca.
 PD WO9801568-A2.
 PF 15-JAN-1998.
 PR 10-JUL-1997; E03670.
 PR 03-MAR-1997; CH-000500.
 PR 10-JUL-1996; CH-001723.
 PA (LONZ) LONZA AG.
 PI Brieden W, Naughton A, Robins K, Shaw N, Tinschert A,
 PI Zimmermann T;
 DR WPI; 98-101063/09.
 PT Tri:fluoro-2-hydroxy-2-methyl-propionic acid enantiomer preparation
 PT - by stereoselective hydrolysis of corresponding racemic amide using
 PT microorganism or derived enzyme, used as drug intermediate
 PS Example 10.2; Page 29; 68pp; German.
 CC W40264-W40275 are peptide fragments of an R-specific amidohydrolase
 CC isolated from Klebsiella oxytoca strain PRS1. This enzyme allows the
 CC microorganism to utilize 3,3,3-trifluoro-2-hydroxy-2-methyl propionamide
 CC as its sole nitrogen source. This amidohydrolase is used in a process for
 CC preparing (R)-isoforms of 3,3,3-trifluoro-2-hydroxy-2-methyl propionamide
 CC which is cheaper than prior art optical resolution of the racemate using
 CC dimethoxy strychnine or (S)-(-)-alpha-methylbenzylamine.
 SQ Sequence 20 AA;

Query Match 6.1%; Score 148; DB 1; Length 20;
 Best Local Similarity 95.0%; Pred. No. 5.37e-03;
 Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 1 MKWLEESIMAKRGVGASRKP 20
 QY 1 MKWLEESIMAKRGVGAGRKP 20

RESULT 5
 ID W40270 standard; Protein; 18 AA.
 AC W40270; 16-JUN-1998 (first entry)
 DT 16-JUN-1998 (first entry)
 DE K. oxytoca R-specific amidohydrolase peptide T8.
 KW R-specific amidohydrolase; hydrolysis; racemate; nitrogen source;
 OS 3,3,3-trifluoro-2-hydroxy-2-methyl propionamide; isoform.
 PN Klebsiella oxytoca.
 PD WO9801568-A2.
 PF 15-JAN-1998.
 PR 10-JUL-1997; E03670.
 PR 03-MAR-1997; CH-000500.
 PR 10-JUL-1996; CH-001723.
 PA (LONZ) LONZA AG.
 PI Brieden W, Naughton A, Robins K, Shaw N, Tinschert A,

```
PI Zimmermann T;
DR WPI; 98-101063/09.
PT Tri:fluoro-2-hydroxy-2-methyl-propionic acid enantiomer preparation
PT - by stereoselective hydrolysis of corresponding racemic amide using
PT microorganism or derived enzyme, used as drug intermediate
PS Example 10.2; Page 29; 68pp; German.
CC W40264-W40275 are peptide fragments of an R-specific amidohydrolase
CC isolated from Klebsiella oxytoca strain PRS1. This enzyme allows the
CC microorganism to utilize 3,3,3-trifluoro-2-hydroxy-2-methyl propionamide
CC as its sole nitrogen source. This amidohydrolase is used in a process for
CC preparing (R)-isoforms of 3,3,3-trifluoro-2-hydroxy-2-methyl propionamide
CC which is cheaper than prior art optical resolution of the racemate using
CC dimethoxy strychnine or (S)-(-)-alpha-methylbenzylamine.
SQ Sequence 18 AA;

Query Match 5.5%; Score 132; DB 1; Length 18;
Best Local Similarity 100.0%; Pred. No. 9.25e-02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 MPFLNPQNGPIMVNGAEK 18
QY 78 MPFLNPQNGPIMVNGAEK 95

RESULT 6
ID W58856 standard; Protein; 575 AA.
AC W58856;
DE 23-JUL-1998 (first entry)
DE C. acidovorans gamma-lactamase protein.
KW Gamma-lactamase; enantiomer; bicyclic lactam; temperature stable;
KW 2-azabicyclo(2.2.1)hept-5-en-3-one; therapeutic agent; antiviral;
KW cardiac vasodilator.
OS Comamonas acidovorans.
PN WO9810075-A1.
PD 12-MAR-1998.
PF 01-SEP-1997; G02344.
PR 03-SEP-1996; GB-018340.
PA (CHIR-) CHIROSCIENCE LTD.
PI Brown RC, Lee CS, Wisdom RA;
DR WPI; 98-193625/17.
DR N-PSDB: V11459.
PT Lactamase enzyme specifically hydrolysing (+)-enantiomer of 2-
PT azabicyclo(2.2.1)hept-5-en-3-one - isolated from Comamonas
PT acidovorans or prepared by recombinant DNA technology, used as, e.g.
PT cardiac vasodilator
PS Claim 5; Page 17-20; 28pp; English.
CC This sequence represents a gamma-lactamase capable of hydrolysing an
CC enantiomer of the bicyclic lactam 2-azabicyclo(2.2.1)hept-5-en-3-one.
CC This enzyme has at least 1 of the following features: greater than 85%
CC retention of activity after incubating at 40 deg. C for 4 hours, or
CC >30% activity after incubating at 60 deg. C for 4 hours, hydrolysis at
CC an initial concentration of 100 g racemic lactam plus 300 ml buffer and
CC proceeding to at least 90 (especially at least 98)% hydrolysis of the
CC (+) lactam with <5% hydrolysis of the (-) lactam. The enzyme is much
CC more temperature-stable than previously identified (+)-gamma-lactamases
CC used for the same purpose. It also enables the bioresolution to be
CC carried out at very high substrate/product concentrations. This
CC enzyme has applications as a therapeutic agent e.g. as an antiviral
CC agent or as a cardiac vasodilator.
SQ Sequence 575 AA;

Query Match 5.5%; Score 132; DB 1; Length 575;
Best Local Similarity 38.0%; Pred. No. 9.25e-02;
Matches 27; Conservative 16; Mismatches 27; Indels 1; Gaps 1;

Db 240 EGARTVPPREHGNCNDIKLSGRSRVFFPYVDGAGLSVGDHLFSQGDGEITFWGPIEMP 299
QY 177 DSINSLTPDHGNGMDVPDIPGSGSIPLYLRPGGRFLFGDAHCACQGDGEICG-TAVEFA 235
Db 300 GWVHMKVSLIK 310
QY 236 SITTIKVDLIK 246
```

```
RESULT 7
ID W40271 standard; Protein; 19 AA.
AC W40271;
DE 16-JUN-1998 (first entry)
DE K. oxytoca R-specific amidohydrolase peptide T9-1.
KW R-specific amidohydrolase; hydrolysis; racemate; nitrogen source;
KW 3,3,3-trifluoro-2-hydroxy-2-methyl propionamide; isoform.
OS Klebsiella oxytoca.
PN WO9801568-A2.
PD 15-JAN-1998.
PF 10-JUL-1997; E03670.
PR 03-MAR-1997; CH-000500.
PR 10-JUL-1996; CH-001723.
PA (LONZ ) LONZA AG.
PI Brieden W, Naughton A, Robins K, Shaw N, Tinschert A,
PI Zimmermann T;
DR WPI; 98-101063/09.
PT Tri:fluoro-2-hydroxy-2-methyl-propionic acid enantiomer preparation
PT - by stereoselective hydrolysis of corresponding racemic amide using
PT microorganism or derived enzyme, used as drug intermediate
PS Example 10.2; Page 29; 68pp; German.
CC W40264-W40275 are peptide fragments of an R-specific amidohydrolase
CC isolated from Klebsiella oxytoca strain PRS1. This enzyme allows the
CC microorganism to utilize 3,3,3-trifluoro-2-hydroxy-2-methyl propionamide
CC as its sole nitrogen source. This amidohydrolase is used in a process for
CC preparing (R)-isoforms of 3,3,3-trifluoro-2-hydroxy-2-methyl propionamide
CC which is cheaper than prior art optical resolution of the racemate using
CC dimethoxy strychnine or (S)-(-)-alpha-methylbenzylamine.
SQ Sequence 19 AA;

Query Match 5.0%; Score 122; DB 1; Length 19;
Best Local Similarity 100.0%; Pred. No. 5.23e-01;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 DAFEGAINSEQDIPSQLLK 19
QY 59 DAFEGAINSEQDIPSQLLK 77

RESULT 8
ID W40273 standard; Protein; 23 AA.
AC W40273;
DE 16-JUN-1998 (first entry)
DE K. oxytoca R-specific amidohydrolase peptide T11.
KW R-specific amidohydrolase; hydrolysis; racemate; nitrogen source;
KW 3,3,3-trifluoro-2-hydroxy-2-methyl propionamide; isoform.
OS Klebsiella oxytoca.
PN WO9801568-A2.
PD 15-JAN-1998.
PF 10-JUL-1997; E03670.
PR 03-MAR-1997; CH-000500.
PR 10-JUL-1996; CH-001723.
PA (LONZ ) LONZA AG.
PI Brieden W, Naughton A, Robins K, Shaw N, Tinschert A,
PI Zimmermann T;
DR WPI; 98-101063/09.
PT Tri:fluoro-2-hydroxy-2-methyl-propionic acid enantiomer preparation
PT - by stereoselective hydrolysis of corresponding racemic amide using
PT microorganism or derived enzyme, used as drug intermediate
PS Example 10.2; Page 29; 68pp; German.
CC W40264-W40275 are peptide fragments of an R-specific amidohydrolase
CC isolated from Klebsiella oxytoca strain PRS1. This enzyme allows the
CC microorganism to utilize 3,3,3-trifluoro-2-hydroxy-2-methyl propionamide
CC as its sole nitrogen source. This amidohydrolase is used in a process for
CC preparing (R)-isoforms of 3,3,3-trifluoro-2-hydroxy-2-methyl propionamide
CC which is cheaper than prior art optical resolution of the racemate using
CC dimethoxy strychnine or (S)-(-)-alpha-methylbenzylamine.
SQ Sequence 23 AA;

Query Match 5.0%; Score 122; DB 1; Length 23;
Best Local Similarity 91.3%; Pred. No. 5.23e-01;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```


Query Match 4.2%; Score 102; DB 1; Length 483;
Best Local Similarity 29.9%; Pred. No. 1.47e+01;

Handwriting practice lines featuring dashed geometric shapes for tracing on a three-line grid. The shapes include a T-shape, an L-shape, a square with a diagonal line, a U-shape, a square, and a chevron shape.

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

protein - protein database search, using Smith-Waterman algorithm

```
Run on: Thu Aug 24 18:55:16 2000; MasPar time 7.21 Seconds
        697.344 Million cell updates/sec
Tabular output not generated.
```

```
Title: >US-09-214-679-2
Description: (1-328) from US09214679.pep
Perfect Score: 2417
Sequence: 1 MKWLEESIMAKRGVAGRKP.....
```

Scoring table: PAM 150
Gap 11

Searched: 152433 seqs, 15329240 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

```
Database:
a-issued
1:5A COMB 2:5B COMB 3:6 COMB 4:PCT COMB 5:backfiles1
```

Statistics: Mean 32.340; Variance 149.359; scale 0.217

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Query			Description	Pred. No.
			Match	Length	DB		
1	102	4.2	483	1	US-08-645-	Sequence 2, Applicatio	8.46e+00
2	100	4.1	483	2	US-08-468-	Sequence 36, Applicati	1.17e+01
3	100	4.1	483	2	US-08-468-	Sequence 32, Applicati	1.17e+01
4	100	4.1	483	1	US-08-468-	Sequence 34, Applicati	1.17e+01
5	100	4.1	483	3	US-08-914-	Sequence 3, Applicatio	1.17e+01
6	100	4.1	483	2	US-08-704-	Sequence 34, Applicati	1.17e+01
7	100	4.1	483	2	US-08-468-	Sequence 36, Applicati	1.17e+01
8	100	4.1	483	3	US-08-890-	Sequence 3, Applicatio	1.17e+01
9	100	4.1	483	4	PCT-US95-1	Sequence 32, Applicati	1.17e+01
10	100	4.1	483	4	PCT-US94-0	Sequence 36, Applicati	1.17e+01
11	100	4.1	483	4	PCT-US94-0	Sequence 32, Applicati	1.17e+01
12	100	4.1	483	4	PCT-US95-1	Sequence 36, Applicati	1.17e+01
13	100	4.1	483	2	US-08-468-	Sequence 32, Applicati	1.17e+01
14	100	4.1	487	2	US-08-468-	Sequence 37, Applicati	1.17e+01
15	100	4.1	487	4	PCT-US94-0	Sequence 37, Applicati	1.17e+01
16	100	4.1	487	2	US-08-468-	Sequence 37, Applicati	1.17e+01
17	100	4.1	487	4	PCT-US95-1	Sequence 37, Applicati	1.17e+01
18	100	4.1	511	4	PCT-US95-1	Sequence 33, Applicati	1.17e+01
19	100	4.1	511	3	US-08-914-	Sequence 2, Applicatio	1.17e+01
20	100	4.1	511	3	US-08-914-	Sequence 4, Applicatio	1.17e+01
21	100	4.1	511	3	US-08-890-	Sequence 4, Applicatio	1.17e+01
22	100	4.1	511	1	US-08-468-	Sequence 35, Applicati	1.17e+01
23	100	4.1	511	1	US-08-645-	Sequence 3, Applicatio	1.17e+01

ALIGNMENTS

```

RESULT          1           STANDARD;      PRT;    483 AA.
XX
XX
AC AC
XX XX
DT DT
XX XX
DE DE
XX XX
Sequence 2, Application US/08645971
Sequence 2, Application US/08645971
Patent No. 5763385
GENERAL INFORMATION:
CC CC
APPLICANT:
CC CC
TITLE OF INVENTION: Modified Alpha-Amylases Having Altered
CC CC                               Calcium Binding Properties
CC CC TITLE OF INVENTION:
CC CC NUMBER OF SEQUENCES:      5
CC CC COMPUTER READABLE FORM:
CC CC MEDIUM TYPE: Floppy disk
CC CC COMPUTER: IBM PC compatible
CC CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC CC SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

```


CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/704,706A
CC FILING DATE: February 20, 1998
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Stone, Christopher
CC REGISTRATION NUMBER: 35,696
CC REFERENCE/DOCKET NUMBER: GC275-3
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (650) 846-7555
CC TELEFAX: (650) 845-6405
CC INFORMATION FOR SEQ ID NO: 34:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 483 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: unknown
CC MOLECULE TYPE: protein
CC SEQUENCE 483 AA; 55211 MW; 1221290 CN;
SQ
Query Match 4.1%; Score 100; DB 2; Length 483;
-Best Local Similarity 29.9%; Pred. No. 1.17e+01;
Matches 23; Conservative 21; Mismatches 29; Indels 4; Gaps 4;
Db 69 OGVTRTKYGTGKELQSA-IKSLHSDINLYGD-VVINHKGADATEDVTAVEVDP-ADR 125
QY 84 QNGPIWNGAEKGDVLAVYIESMLPRGVDPYGCAMIPHFGGLTGT-DLTAMLNDPLPEK 142
Db 126 NRVISGEHLIKAWTHF 142
QY 143 VRMIKLDSEKYYWSKRH 159
RESULT 7
ID US-08-468-698-36 STANDARD; PRT; 483 AA.
XX XXXXXX
AC XXXXXX
XX
DT
XX
DE
XX
XX
Sequence 36, Application US/08468698
Sequence 36, Application US/08468698
Patent No. 5849549
GENERAL INFORMATION:
CC APPLICANT: Antrim, Richard L.
CC APPLICANT: Barnett, Christopher
CC APPLICANT: Mitchinson, Colin
CC APPLICANT: Power, Scott D.
CC APPLICANT: Requaadt, Carol
CC APPLICANT: Solheim, Leif P.
CC TITLE OF INVENTION: Oxidatively Stable Alpha-Amylase
CC NUMBER OF SEQUENCES: 68
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Genencor International, Inc.
CC STREET: 180 Kimball Way
CC CITY: South San Francisco
CC STATE: CA
CC COUNTRY: USA
CC ZIP: 94080
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: 3.5 inch, 1.44 Mb
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/468,698
CC FILING DATE: 06-JUN-95

CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/194,664
CC FILING DATE: 10-FEB-94
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/016,395
CC FILING DATE: 11-FEB-93
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Stone, Christopher L.
CC REGISTRATION NUMBER: 35,696
CC REFERENCE/DOCKET NUMBER: GC220D2
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (415) 742-7555
CC TELEFAX: (415) 742-7217
CC INFORMATION FOR SEQ ID NO: 36:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 483 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 483 AA; 55181 MW; 1222078 CN;
SQ
Query Match 4.1%; Score 100; DB 2; Length 483;
-Best Local Similarity 29.9%; Pred. No. 1.17e+01;
Matches 23; Conservative 21; Mismatches 29; Indels 4; Gaps 4;
Db 69 OGVTRTKYGTGKELQSA-IKSLHSDINLYGD-VVINHKGADATEDVTAVEVDP-ADR 125
QY 84 QNGPIWNGAEKGDVLAVYIESMLPRGVDPYGCAMIPHFGGLTGT-DLTAMLNDPLPEK 142
Db 126 NRVISGEHLIKAWTHF 142
QY 143 VRMIKLDSEKYYWSKRH 159
RESULT 8
ID US-08-890-383-3 STANDARD; PRT; 483 AA.
XX XXXXXX
AC XXXXXX
XX
DT
XX
DE
XX
XX
Sequence 3, Application US/08890383
Sequence 3, Application US/08890383
Patent No. 6008026
GENERAL INFORMATION:
CC APPLICANT: Anthony G. Day
CC TITLE OF INVENTION: MUTANT ALPHA-AMYLASE HAVING INTRODUCED
CC TITLE OF INVENTION: THEREIN A DISULFIDE BOND
CC NUMBER OF SEQUENCES: 11
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Genencor International, Inc.
CC STREET: 925 Page Mill Road
CC CITY: Palo Alto
CC STATE: CA
CC COUNTRY: USA
CC ZIP: 94304-1013
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette
CC COMPUTER: IBM Compatible
CC OPERATING SYSTEM: DOS
CC SOFTWARE: FastSeq for Windows Version 2.0
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/890,383
CC FILING DATE: To Be Assigned
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Stone, Christopher
CC REGISTRATION NUMBER: 35,696
CC REFERENCE/DOCKET NUMBER: GC377
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (650) 846-7555

DT XX Sequence 32, Application PC/TUS9401553A
DE XX Sequence 32, Application PC/TUS9401553A
CC XX GENERAL INFORMATION:
CC APPLICANT: GENENCOR INTERNATIONAL, INC.
CC TITLE OF INVENTION: Oxidatively Stable Alpha-Amylase
CC NUMBER OF SEQUENCES: 68
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Genencor International, Inc.
CC STREET: 180 Kimball Way
CC CITY: South San Francisco
CC STATE: CA
CC COUNTRY: USA
CC ZIP: 94080
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US94/01553A
CC FILING DATE:
CC CLASSIFICATION:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Horn, Margaret A.
CC REGISTRATION NUMBER: 33,401
CC REFERENCE/DOCKET NUMBER: GC220-2
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (415) 742-7536
CC TELEFAX: (415) 742-7217
CC INFORMATION FOR SEQ ID NO: 32:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 483 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 483 AA; 55211 MW; 1221290 CN;
Query Match 4.1%; Score 100; DB 4; Length 483;
Best Local Similarity 29.9%; Pred. No. 1.17e+01;
Matches 23; Conservative 21; Mismatches 29; Indels 4; Gaps 4;
Db 69 QKGTVRTKYGKGELOSA-IKLSHRDINVYGD-VVINHGKGDATEDVTAIVEDP-ADR 125
QY 84 QNGPIWNGAEKGDVLAVIESMLPRGVDYPYICAMIPHFGGLTGT-DITAMNDPLPEK 142
Db 126 NRVISGEHLIKAWTHFH 142
QY 143 VRMIKLDSEKVVYWSRRH 159
RESULT 12
ID PCT-US95-10426-36 STANDARD; PRT: 483 AA.
XX xxxxxx
DT
DE
XX Sequence 36, Application PC/TUS9510426
XX Sequence 36, Application PC/TUS9510426
CC GENERAL INFORMATION:
CC APPLICANT: GENENCOR INTERNATIONAL, INC.
CC TITLE OF INVENTION: An Improved Cleaning Composition
CC NUMBER OF SEQUENCES: 68
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Genencor International
CC STREET: 180 Kimball Way
CC CITY: South San Francisco
CC STATE: CA

CC COUNTRY: USA
CC ZIP: 94080
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US95/10426
CC FILING DATE:
CC CLASSIFICATION:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: STONE, Christopher L.
CC REGISTRATION NUMBER: 33,401
CC REFERENCE/DOCKET NUMBER: GC220-3
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (415) 742-7536
CC TELEFAX: (415) 742-7217
CC INFORMATION FOR SEQ ID NO: 36:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 483 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 483 AA; 55181 MW; 1222078 CN;
Query Match 4.1%; Score 100; DB 4; Length 483;
Best Local Similarity 29.9%; Pred. No. 1.17e+01;
Matches 23; Conservative 21; Mismatches 29; Indels 4; Gaps 4;
Db 69 QKGTVRTKYGKGELOSA-IKLSHRDINVYGD-VVINHGKGDATEDVTAIVEDP-ADR 125
QY 84 QNGPIWNGAEKGDVLAVIESMLPRGVDYPYICAMIPHFGGLTGT-DITAMNDPLPEK 142
Db 126 NRVISGEHLIKAWTHFH 142
QY 143 VRMIKLDSEKVVYWSRRH 159
RESULT 13
ID US-08-468-220-32 STANDARD; PRT: 483 AA.
XX xxxxxx
DT
DE
XX Sequence 32, Application US/08468220
XX Sequence 32, Application US/08468220
CC Patent No. 5824532
CC GENERAL INFORMATION:
CC APPLICANT: Antrim, Richard L.
CC APPLICANT: Barnett, Christopher
CC APPLICANT: Mitchinson, Colin
CC APPLICANT: Power, Scott D.
CC APPLICANT: Requaft, Carol
CC APPLICANT: Solheim, Leif P.
CC TITLE OF INVENTION: Oxidatively Stable Alpha-Amylase
CC NUMBER OF SEQUENCES: 68
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Genencor International, Inc.
CC STREET: 180 Kimball Way
CC CITY: South San Francisco
CC STATE: CA
CC COUNTRY: USA
CC ZIP: 94080
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: 3.5 inch, 1.44 Mb
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:

CC INFORMATION FOR SEQ ID NO: 37:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 487 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SQ SEQUENCE 487 AA; 55495 MW; 1241380 CN;

Query Match 4.1%; Score 100; DB 4; Length 487;
Best Local Similarity 29.9%; Pred. No. 1.17e+01;
Matches 23; Conservative 21; Mismatches 29; Indels 4; Gaps 4;
Db 73 QKGTVRTKYGTELQSA-IKSLHSRDINVYGD-VVINHKGGADATEDVTAVEVDP-ADR 129
QY 84 QNGPINVNGAEGDVLAVYIESMLPRGVDPYGICAMIPHFGGLTGT-DLTAMINDPLPEK 142
Db 130 NRVISGEHLIKAWTFH 146
QY 143 VRMIKLDSEKVIWVKRH 159

Search completed: Thu Aug 24 18:55:43 2000
Job time : 27 secs.


```

RESULT      2
ENTRY       G72414      #type complete
TITLE       hypothetical protein TM0119 - Thermotoga maritima (strain
             MSB8)
ORGANISM    #formal_name Thermotoga maritima
DATE        11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change
ACCESSIONS  G72414
REFERENCE   A72200
#authors    Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson,
             R.J.; Haft, D.H.; Hickey, E.K.; Peterson, J.D.; Nelson,
             W.C.; Ketchum, K.A.; McDonald, L.; Otterback, T.R.; Malek,
             J.A.; Linher, K.D.; Garrett, M.M.; Stewart, A.M.; Cotton,
             M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
             Heidelberg, J.; Sutton, G.G.; Fleischmann, R.D.; White, O.;
             Salzberg, S.L.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
#journal    Nature (1999) 399:323-329
#title      Evidence for lateral gene transfer between Archaea and
             Bacteria from genome sequence of Thermotoga maritima.
#cross-references MUID:99287316
#accession   G72414
#status      preliminary
#molecule_type DNA
#residues    1-285 #label ARN
#cross-references GB:AE001698; GB:AE000512; NID:g4980609; PID:g4980611;
             TIGR:TM0119
#experimental_source strain MSB8

GENETICS
#gene        TM0119
SUMMARY      #length 285 #molecular-weight 31196 #checksum 8059
             #length 285; Score 266; DB 2; Length 285;
             Best Local Similarity 30.8%; Pred. No. 3.15e-27;
             Matches 52; Conservative 40; Mismatches 72; Indels 5; Gaps 5;

Db 117 IPIHMGVIGVAPGEVPTGTARRHGGNMDTKET-TEYTVHLFPVQEGNALALGDVH 175
QY 161 LPYKPHITGLSVSPIDISINSLDHNHGGNMDVPDIPGSGITYPL-VRAPGGRFLFIGDAH 219
Db 176 ATMGDGEVCVACEPAKVAVVVDYSKE-EIKVPVVFENDAYIIVSLPDIIEEALKVETR 234
QY 220 ACQGDGEICGTAVEFASITTIKVDLIKKNQLSWPRMENAIMSGSARPLEDATRIAYR 279
Db 235 ETV-WFIQRRKTIPTDAYMLASVDVGVISOLVNPAKTAKARIPKIYF 282
QY 280 DLIYWLVEDFGFEOW-DAYMLLSQCGKVRLLGNMVDPKYTVGMLNKNLL 327

RESULT      3
ENTRY       C72750      #type complete
TITLE       probable acetamidase APE0528 - Aeropyrum pernix (strain K1)
ORGANISM    #formal_name Aeropyrum pernix
DATE        20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change
ACCESSIONS  C72750
REFERENCE   A72450
#authors    Kawarabayasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.;
             Haikawa, Y.; Jin-no, K.; Takahashi, M.; Sekine, M.; Baba,
             S.; Ankal, A.; Kosugi, H.; Hosoyama, A.; Fukui, S.; Nagai,
             Y.; Nishijima, K.; Nakazawa, H.; Takamiya, M.; Masuda, S.;
             Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.;
             Kushida, N.; Oguchi, A.; Aoki, K.; Kubota, K.; Nakamura,
             Y.; Nomura, N.; Sako, Y.; Kikuchi, H.
#journal    DNA Res. (1999) 6:83-101
#title      Complete genome sequence of an aerobic hyper-thermophilic
             Crenarchaeon, Aeropyrum pernix K1.
#cross-references MUID:99310339
#accession   C72750
#status      preliminary
#molecule_type DNA
#residues    1-377 #label KAW

```

```

#cross-references DBJ:AP000059; NID:g5103911; PIDN:BAA79495.1;
             PID:d1043281; PID:g5104179
#experimental_source strain K1

GENETICS
#gene        APE0528
SUMMARY      #length 377 #molecular-weight 41052 #checksum 5419
             Query Match 8.2%; Score 197; DB 2; Length 377;
             Best Local Similarity 34.2%; Pred. No. 4.84e-15;
             Matches 40; Conservative 27; Mismatches 47; Indels 3; Gaps 3;

Db 146 LGVMGVARSKELLKEIKERDLKRGFVLPPPEGAVPPREPVASGLRTIPPRENGG 205
QY 131 LTAMLNDPLPEKVRMIKLDSEKVVWSKRHTLPYKPHIGTSLVSP-EIDSINSLTPDNGHG 189
Db 206 NLDVRHFSPPSGKIFYPVPEGA-LFSVGDHAHQDGVCGTAIMGATATLRFQVI 261
QY 190 NMDVPDIPGSGITYPLVRAPGGRFL-IGDAHACQGDGEICGTAVEFASITTIKVDLI 245

RESULT      4
ENTRY       E71097      #type complete
TITLE       hypothetical protein PH1041 - Pyrococcus horikoshii
ORGANISM    #formal_name Pyrococcus horikoshii
DATE        14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change
ACCESSIONS  E71097
REFERENCE   A71000
#authors    Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.;
             Hino, Y.; Yamamoto, S.; Sekine, M.; Baba, S.; Kosugi, H.;
             Hosoyama, A.; Nagai, Y.; Sakai, M.; Ogura, K.; Otsuka, R.;
             Nakazawa, H.; Takamiya, M.; Ohfuku, Y.; Funahashi, T.;
             Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kushida, N.; Oguchi,
             A.; Aoki, K.; Yoshizawa, T.; Nakamura, Y.; Robb, F.T.;
             Horikoshi, K.; Masuchi, Y.; Shizuya, H.; Kikuchi, H.
#journal    DNA Res. (1998) 5:55-76
#title      Complete sequence and gene organization of the genome of a
             hyper-thermophilic archaeobacterium, Pyrococcus horikoshii
             OT3.
#cross-references MUID:98344137
#accession   E71097
#status      preliminary; nucleic acid sequence not shown;
             #molecule_type DNA
             translation not shown
#residues    1-389 #label KAW
#cross-references GB:AP000004; NID:g3236131; PID:d1031082; PID:g3257456
#experimental_source strain OT3
#note        this accession replaces an interim accession for a
             sequence replaced by GenBank

GENETICS
#gene        PH1041
SUMMARY      #length 389 #molecular-weight 42005 #checksum 7333
             Query Match 6.7%; Score 163; DB 2; Length 389;
             Best Local Similarity 38.7%; Pred. No. 1.91e-09;
             Matches 24; Conservative 17; Mismatches 21; Indels 0; Gaps 0;

Db 206 GHLDVDVRGCAIVIAPVKVGGVYAGDAHAMGDGEVAGHTTDTVTAETVTVEVIKGL 265
QY 189 GNMDVPDIPGSGITYPLVRAPGGRFLFIGDAHACQGDGEICGTAVEFASITTIKVDLINW 248
Db 266 NL 267
QY 249 QL 250

RESULT      5
ENTRY       S74213      #type complete
TITLE       formamidase (EC 3.5.1.49) A - Methylophilus methylotrophus
ORGANISM    #formal_name Methylophilus methylotrophus
DATE        04-Dec-1997 #sequence_revision 12-Dec-1997 #text_change
ACCESSIONS  S74213; S78115

```

```
REFERENCE S74213
#authors Wyborn, N.R.; Mills, J.; Williams, S.G.; Jones, C.W.
#journal Eur. J. Biochem. (1996) 240:314-322
#title Molecular characterisation of formamidase from Methylophilus
        methylophilus.
#cross-references MUID:96439058
#accession S74213
#molecule_type DNA
##residues 1-407 ##label WYB
##cross-references EMBL:X99632; NID:g1480104; PID:e256826; PID:g1480105
##experimental_source strain NCIB 10515
#accession S78115
#molecule_type protein
##residues 1-5;44-47;172-176;211-215 ##label WYC
GENETICS
#gene fmdA
#keywords hydrolase
SUMMARY
#length 407 #molecular-weight 44466 #checksum 2787
Query Match 6.5%; Score 157; DB 2; Length 407;
Best Local Similarity 45.3%; Pred. No. 1.69e-08;
Matches 34; Conservative 14; Mismatches 20; Indels 7; Gaps 6;
Db 210 EGARTVPPRHGGNCDIKNLTKSRVFPVYVKD-GG-LSMGDLHFSQGDGEITFCG-AI 266
QY 177 DSINSLTPDNHGGNMDVPDIPGSGITY-PL-VRAPGGRGLFTGDAHACQGDGEI--CGTAV 232
Db 267 EMAGYLDIKVGLIKD 281
QY 233 EFASITTIKVLDLIKN 247
RESULT 6
ENTRY A47696 #type complete
TITLE acetamidase - Mycobacterium smegmatis
ORGANISM #formal_name Mycobacterium smegmatis
DATE 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change
ACCESSIONS A47696
REFERENCE A47696
#authors Mahenthiralingam, E.; Draper, P.; Davils, E.O.; Colston, M.J.
#journal J. Gen. Microbiol. (1993) 139:575-583
#title Cloning and sequencing of the gene which encodes the highly
        inducible acetamidase of Mycobacterium smegmatis.
#cross-references MUID:93232774
#contents NCTC 8159
#accession A47696
#status preliminary
#molecule_type DNA; protein
##residues 1-406 ##label MAH
##cross-references GB:X57175; GB:S59037; NID:g312080; PID:g312081
##note sequence inconsistent with the nucleotide translation
        sequence extracted from NCBI backbone (NCBIN:129965,
        NCBIP:129966)
SUMMARY #length 406 #molecular-weight 43964 #checksum 926
Query Match 6.1%; Score 147; DB 2; Length 406;
Best Local Similarity 43.8%; Pred. No. 5.95e-07;
Matches 32; Conservative 14; Mismatches 22; Indels 5; Gaps 4;
Db 218 DGARTVPPRHGGNMDIKNFTGRGRIFFP-VFVEGAMSLGGDLHFSQGDGEINFCG-AIE 275
QY 177 DSINSLTPDNHGGNMDVPDIPGSGS-ITYPLVRAPGGRGLFTGDAHACQGDGEI--CGTAVE 233
Db 276 MGGFIDMHVDLIK 288
QY 234 FASITTIKVLDLIK 246
RESULT 7
ENTRY T04712 #type complete
TITLE probable formamidase (EC 3.5.1.49) F19F18.40 - Arabidopsis
        thaliana
```

```
ORGANISM #formal_name Arabidopsis thaliana #common_name mouse-ear
        cress
DATE 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change
        11-Jun-1999
ACCESSIONS T04712
REFERENCE Z15382
#authors Bevan, M.; Rose, M.; Hempel, S.; Entian, K.D.; Bancroft, I.;
        Mewes, H.W.; Mayer, K.F.X.; Schueller, C.
#submission submitted to the Protein Sequence Database, March 1999
#accession T04712
#molecule_type DNA
##residues 1-432 ##label BEV
##cross-references EMBL:AL035605
##experimental_source cultivar Columbia; BAC clone F19F18
GENETICS
#map_position 4
#introns 73/3; 153/1; 196/3; 275/2; 356/3; 392/3
#note F19F18.40
#keywords hydrolase
SUMMARY #length 432 #molecular-weight 47566 #checksum 5561
Query Match 5.6%; Score 135; DB 2; Length 432;
Best Local Similarity 35.6%; Pred. No. 3.72e-05;
Matches 26; Conservative 19; Mismatches 25; Indels 3; Gaps 2;
Db 210 EAARTIPGRENGNODIKNLRSKYLPLVFVEGANLSTGDMHFSQGDGEISFCG-AIEM 268
QY 177 DSINSLTPDNHGGNMDVPDIPGSGITYPLVRAPGGRGLFTGDAHACQGDGEI--CGTAVEF 234
Db 269 SGFLEKCEIIRN 281
QY 235 ASITTIKVLDLIKN 247
RESULT 8
ENTRY B70434 #type complete
TITLE polysialic acid capsule expression protein - Aquifex aeolicus
ORGANISM #formal_name Aquifex aeolicus
DATE 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change
        29-Sep-1999
ACCESSIONS B70434
REFERENCE A70300
#authors Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.;
        Lennox, A.L.; Graham, D.E.; Overbeek, R.; Snead, M.A.;
        Keller, M.; AuJay, M.; Huber, R.; Feldman, R.A.; Short,
        J.M.; Olson, G.J.; Swanson, R.V.
#journal Nature (1998) 392:353-358
#title The complete genome of the hyperthermophilic bacterium
        Aquifex aeolicus.
#cross-references MUID:98196666
#accession B70434
#status preliminary; nucleic acid sequence not shown;
        translation not shown
#molecule_type DNA
##residues 1-322 ##label AOF
##cross-references GB:AE000745; NID:g2983907; PIDN:AAC07460.1;
        PID:g2983910; GB:AE000657
##experimental_source strain VF5
GENETICS
#gene kpsF
#classification #superfamily probable ATP-binding protein gutQ; CBS homology
FEATURE
208-256 #domain CBS homology #label CBS1\
273-321 #domain CBS homology #label CBS2
SUMMARY #length 322 #molecular-weight 35114 #checksum 6980
Query Match 4.3%; Score 103; DB 1; Length 322;
Best Local Similarity 25.2%; Pred. No. 8.78e-01;
Matches 28; Conservative 35; Mismatches 41; Indels 7; Gaps 7;
Db 83 GDMGLLDSKDALI-AISNSGSGSTEVLVLYQAKALNIPVIGITGNEKSS-LAKYSDVLIK 140
QY 50 GDRITVDTRDAFEAGINSEQDIPSQLKMPFLNPONGPIM-VNGAEKGDVLAITYESMLP 108
```

```

Db 141 IPVDR-EACP-F-NLAPTSSVTTLALGDAIATMLKLGFSQEDF-AKRH 187
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 109 RGVDPYGCAMPHFGGLTGTDLTAMLNDPLPEKVRMIKLDSEKVVYSKRH 159

RESULT 9
ENTRY #type complete
TITLE hypothetical protein rv3025c - Mycobacterium tuberculosis
ORGANISM (strain H37RV)
DATE 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
ACCESSION D70858
REFERENCE A70500
#authors Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry III, C.E.; Tekaia, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrall, B.G.
#journal Nature (1998) 393:537-544
#title Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.
#cross-references MUID:98295987
#accession D70858
#status preliminary; nucleic acid sequence not shown; translation not shown
#molecule_type DNA
##residues 1-393 ##label COL
##cross-references GB:AL021287; GB:AL123456; NID:g3261508; PID:e1237771;
##experimental_source strain H37RV
GENETICS
#gene RV3025c
CLASSIFICATION #superfamily nitrogen fixation protein nifs
FEATURE 329
#active_site Cys (cysteine persulfide intermediate)
#status predicted
SUMMARY #length 393 #molecular-weight 40947 #checksum 4169
Query Match 4.3%; Score 103; DB 2; Length 393;
Best Local Similarity 25.8%; Pred. No. 8.78e-01;
Matches 17; Conservative 20; Mismatches 25; Indels 4; Gaps 4;

Db 73 NLAVGIYWARDAPHRRIIVTVEHHAVLDSVNLV-EHEGAHVTLPTAADGSVSA 131
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 147 KLDSEKVVYSKRHLPYKPHIGTLVSPE-I-DSINSLTPDNHGGNMD-VPDIPGCSITY 203
Db 132 TALREA 137
|| :
QY 204 PLVRAP 209

RESULT 10
ENTRY #type complete
TITLE alkylglycerone-phosphate synthase (EC 2.5.1.26) - Caenorhabditis elegans
ALTERNATE_NAMES alkyldihydroxyacetonephosphate synthase
ORGANISM #formal_name Caenorhabditis elegans
DATE 20-Feb-1998 #sequence_revision 13-Mar-1998 #text_change
ACCESSION JC5829
REFERENCE 07-May-1999
#authors de Vet, E.C.J.M.; Prinsen, H.C.M.T.; van den Bosch, H.
#journal Biochem. Biophys. Res. Commun. (1998) 242:277-281
#title Nucleotide sequence of a cDNA clone encoding a Caenorhabditis elegans homolog of mammalian alkyl-dihydroxyacetonephosphate synthase: Evolutionary
switching of peroxisomal targeting signals.
#cross-references MUID:98113342
#accession JC5829
#molecule_type mRNA
##residues 1-597 ##label DEV
##cross-references GB:AJ002686; NID:g2821944; PID:e1248248; PID:g2821945
COMMENT This enzyme catalyzes the exchange of the acyl chain for a long chain fatty alcohol.
KEYWORDS transferase
SUMMARY #length 597 #molecular-weight 66559 #checksum 3790
Query Match 4.3%; Score 104; DB 2; Length 597;
Best Local Similarity 28.2%; Pred. No. 6.58e-01;
Matches 24; Conservative 25; Mismatches 29; Indels 7; Gaps 5;

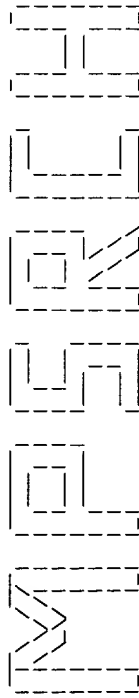
Db 159 CALIP-IGS-CTSVYNALDTPETKRAVISMDMLDKILMDRENTLCRAQAGIVGOS 215
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 117 CAMIPHFGGLTGTDLTAMLNDPLPEKVRMIKLD-S--EKVYWSKRHTLPYKPHIGTLVS 173
Db 216 LERQLNKGFTCGHEPDSIERFTLG 240
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 174 PEID-SINSLTPDNHGGNMDVDPDG 197

RESULT 11
ENTRY #type complete
TITLE vitamin B12 transport permease protein btuC precursor - Escherichia coli
ORGANISM #formal_name Escherichia coli
DATE 30-Jun-1988 #sequence_revision 05-Dec-1997 #text_change
ACCESSION G64929; A24498; S04777
REFERENCE A64720
#authors Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.
#journal Science (1997) 277:1453-1462
#title The complete genome sequence of Escherichia coli K-12.
#cross-references MUID:97426617
#accession G64929
#status nucleic acid sequence not shown; translation not shown
#molecule_type DNA
##residues 1-326 ##label BLAT
##cross-references GB:AE000266; GB:U00096; NID:g1787997;
##experimental_source strain K-12, substrain MG1655
REFERENCE A24498
#authors Friedrich, M.J.; DeVeaux, L.C.; Kadner, R.J.
#journal J. Bacteriol. (1986) 167:928-934
#title Nucleotide sequence of the btuCD genes involved in vitamin B12 transport in Escherichia coli and homology with components of periplasmic-binding-protein-dependent transport systems.
#cross-references MUID:86304184
#accession A24498
#molecule_type DNA
##residues 1-110, 'LN', 113-122, 'R', 124-285, 'DCCWPIL' ##label PRI
#note this sequence has been revised in reference S04777
REFERENCE S04777
#authors Rioux, C.R.; Kadner, R.J.
#journal Mol. Gen. Genet. (1989) 217:301-308
#title Vitamin B(12) transport in Escherichia coli K12 does not require the btuE gene of the btuCED operon.
#cross-references MUID:89364713
#accession S04777
#status nucleic acid sequence not shown
#molecule_type DNA
##residues 280-326 ##label RIO
COMMENT This membrane-associated protein is required for vitamin B12 transport across the cytoplasmic membrane; however, its exact role is not clear.

```

[illegible][illegible]

Search completed: Thu Aug 24 18:54:13 2000
Job time : 54 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Thu Aug 24 18:51:08 2000; MasPar time 11.48 seconds
Tabular output not generated. 885.716 Million cell updates/sec

Title: >US-09-214-679-2
Description: (1-328) from US09214679.ppep
Perfect Score: 2417
Sequence: 1 MKWLEESIMAKRGVAGRKP.....GNMVDPKYTCAMLNKLLV 328
Scoring table: PAM 150
Gap 11
Searched: 85661 seqs, 30989116 residues
Post-processing: Minimum Match 0%
Listing first 45 summaries
Database: swiss-prot38
1:swissprot

Statistics: Mean 49.160; Variance 85.522; scale 0.575

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	157	6.5	407	1	FMDA_METME FORMAMIDASE (EC 3.5.1.1)	6.09e-10
2	147	6.1	405	1	AMDA_MYCSM ACETAMIDASE (EC 3.5.1.1)	3.27e-08
3	125	5.2	375	1	FTSZ_HALSA CELL DIVISION PROTEIN	1.33e-04
4	110	4.6	498	1	MEFA_MOUSE MYOCYTE-SPECIFIC ENHANCER	2.45e-02
5	103	4.3	322	1	YF46_AQUAE VITAMIN B12 TRANSPORT	2.40e-01
6	101	4.2	326	1	BTUC_ECOLI HYPOHETICAL 41.2 KDA	4.50e-01
7	102	4.2	512	1	AMP5_ECOLI ALPHA-AMYLASE PRECURSOR	4.29e-01
8	101	4.2	810	1	UNKNOW PROTEIN FROM 2	3.29e-01
9	102	4.2	924	1	104K_THEPA 104 KDA MICRONEME-RHOP	4.50e-01
10	101	4.2	4969	1	RYNC_RABIT RYANODINE RECEPTOR, CA	8.38e-01
11	99	4.1	181	1	ARF_CRYNE ADP-RIBOSYLATION FACTO	1.14e+00
12	98	4.1	380	1	FTSZ_MYCPN CELL DIVISION PROTEIN	1.14e+00
13	98	4.1	396	1	CHSB_IPOPU CHALCONE SYNTHASE B (E	8.38e-01
14	99	4.1	397	1	CAPB_BACAN CAPB PROTEIN	1.14e+00
15	98	4.1	460	1	FD3C_RICCO OMEGA-3 FATTY ACID DES	1.14e+00
16	98	4.1	632	1	RCAC_FREDI RCAC PROTEIN	8.38e-01
17	99	4.1	5032	1	RYNR_HUMAN RYANODINE RECEPTOR, SK	8.38e-01
18	99	4.1	5035	1	RYNR_PIG RYANODINE RECEPTOR, SK	2.08e+00
19	96	4.0	135	1	PILG_PSEAE PILG PROTEIN	1.54e+00
20	97	4.0	140	1	JANE_DROME SEX-REGULATED PROTEIN	2.08e+00
21	96	4.0	180	1	ARFL_ARATH ADP-RIBOSYLATION FACTO	2.08e+00
22	96	4.0	180	1	ARF_ORYSA ADP-RIBOSYLATION FACTO	2.08e+00
23	96	4.0	180	1	ARF_MAIZE ADP-RIBOSYLATION FACTO	2.08e+00

24	96	4.0	180	1	ARFL_CHLRE ADP-RIBOSYLATION FACTO	2.08e+00
25	97	4.0	500	1	ABFA_BACSU ALPHA-L-ARABINOFURANOS	1.54e+00
26	96	4.0	1249	1	TTP2_HUMAN TRIPEPTIDYL-PEPTIDASE	2.08e+00
27	97	4.0	1853	1	CIPA_CLOTM CELLULOSONAL SCAFFOLDI	1.54e+00
28	95	3.9	340	1	CARP_POLTU POLYPOPEPTIN (EC 3.4	2.81e+00
29	95	3.9	396	1	CHSA_IPOPU CHALCONE SYNTHASE A (E	2.81e+00
30	95	3.9	417	1	IE63_HSVSA 52 KDA IMMEDIATE-EARLY	2.81e+00
31	95	3.9	446	1	FD3C_ARATH OMEGA-3 FATTY ACID DES	2.81e+00
32	94	3.9	498	1	YHJJ_ECOLI PROTEIN YHJJ PRECURSOR	3.78e+00
33	95	3.9	605	1	VE1_HPV09 REPLICATION PROTEIN E1	2.81e+00
34	95	3.9	1262	1	TTP2_MOUSE TRIPEPTIDYL-PEPTIDASE	2.81e+00
35	95	3.9	3390	1	POLG_DEN3 GENOME POLYPROTEIN (CO	2.81e+00
36	92	3.8	180	1	ARFL_DAUCA ADP-RIBOSYLATION FACTO	6.77e+00
37	93	3.8	182	1	ARF_AJECA ADP-RIBOSYLATION FACTO	5.06e+00
38	92	3.8	182	1	FTNL_HAEIN FERRITIN LIKE PROTEIN	6.77e+00
39	93	3.8	363	1	CHSB_IPOCO CHALCONE SYNTHASE B (E	5.06e+00
40	92	3.8	373	1	YN28_YEAST HYPOHETICAL 41.2 KDA	6.77e+00
41	93	3.8	560	1	MIS_HUMAN MUELLERIAN INHIBITING	5.06e+00
42	93	3.8	772	1	CIPB_CLOTM CELLULOSONAL SCAFFOLDI	5.06e+00
43	93	3.8	1010	1	CONT_CHICK CONTACTIN PRECURSOR (N	5.06e+00
44	92	3.8	1068	1	SPS_MAIZE SUCROSE-PHOSPHATE SYNT	6.77e+00
45	92	3.8	1105	1	YEGE_ECOLI HYPOHETICAL 123.9 KDA	6.77e+00

ALIGNMENTS

RESULT 1
ID FMDA_METME STANDARD; PRT; 407 AA.
AC Q50228;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE FORMAMIDASE (EC 3.5.1.49) (FORMAMIDE AMIDOHYDROLASE).
GN FMDA.
OS Methylophilus methylotrophus (Bacterium W3A1).
OC Bacteria; Proteobacteria; beta subdivision; Methylophilus group;
OC Methylophilus.
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX STRAIN=NCIB 10515;
RX MEDLINE: 96439058.
RA Wyborn N.R., Mills J., Williams S.G., Jones C.W.;
RT "Molecular characterisation of formamidase from Methylophilus
methylophilus.";
RL Eur. J. Biochem. 240:314-322(1996).
RN [2]
RP CHARACTERIZATION, AND SEQUENCE OF 1-5.
RC STRAIN=NCIB 10515;
RA Wyborn N.R., Scherr D.J., Jones C.W.;
RT "Purification, properties and heterologous expression of formamidase
from Methylophilus methylotrophus.";
RL Microbiology 140:191-195(1994).
CC -!- FUNCTION: HYDROLYSE FORMAMIDE WITH THE PRODUCTION OF AMMONIA WHICH
CAN BE USED AS A SOURCE OF NITROGEN FOR GROWTH. ALSO ACTS, MORE
SLOWLY, ON ACETAMIDE, PROPANAMIDE AND BUTANAMIDE.
CC -!- CATALYTIC ACTIVITY: FORMAMIDE + H(2)O = FORMATE + NH(3).
CC -!- SUBUNIT: HOMOTRIMER.
CC -!- MASS SPECTROMETRY: MW=44481; MW_ERR=26; METHOD=ELECTROSPRAY.
CC -!- SIMILARITY: STRONG, TO M.SMEGMATIS ACETAMIDASE.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

DR EMBL: X99632; CAA67953.1;
KW Hydrolyase.
SQ SEQUENCE 407 AA; 44466 MW; C5E8888BA4C44A70 CRC64;
Query Match 6.5%; Score 157; DB 1; Length 407;

```

Best Local Similarity 45.3%; Pred. No. 6.09e-10;
Matches 34; Conservative 14; Mismatches 20; Indels 7; Gaps 6;

Db 210 EGARTVPPRDHGGNGDINKLTKGSRVPPVVKD-GG-LSMGDLHFSOGDGEITFCG-AI 266
QY 177 DSINSLTPDNHGGNMDVPDIPGSGITY-PL-VRAPEGRLFIGDAHACQDGEI--CGTAV 232
Db 267 EMAGYLDIKVGLIKD 281
QY 233 EFASITTIKVLDLIK 247

RESULT 2
ID AMDA_MYCSM STANDARD; PRT; 405 AA.
AC Q07838;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE ACETAMIDASE (EC 3.5.1.-).
OS Mycobacterium smegmatis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Mycobacterium.
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-23; 81-100 AND 180-194.
RC STRAIN=NCTC 8159;
RX MEDLINE; 93232774.
RA Mahenthalingam E., Draper P., Davis E.O., Colston M.J.;
RT "Cloning and sequencing of the gene which encodes the highly
RT inducible acetamidase of Mycobacterium smegmatis.";
RL J. Gen. Microbiol. 139:575-583(1993).
CC -1- FUNCTION: ALLOWS ACETAMIDE TO BE USED AS A SOLE CARBON OR NITROGEN
CC SOURCE.
CC -1- CATALYTIC ACTIVITY: FORMAMIDE + H(2)O = FORMATE + NH(3).
CC -1- SIMILARITY: STRONG. TO M.METHYLOTROPHUS FORMAMIDASE.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X57175; CAA0462.1; -.
CC KW Hydrolase.
CC FT INIT_MET 0 0 POTENTIAL.
CC SQ SEQUENCE 405 AA; 43833 MW; 05E2E359210E0BF2 CRC64;

Query Match 6.1%; Score 147; DB 1; Length 405;
Best Local Similarity 43.8%; Pred. No. 3.27e-08;
Matches 32; Conservative 14; Mismatches 22; Indels 5; Gaps 4;

Db 217 DGARTVPPRENGNHDIKNTRGSRIFYP-VFVEGAMLSGGDLHFSOGDGEINFCG-AIE 274
QY 177 DSINSLTPDNHGGNMDVPDIPGSGS-ITVPLVRAPEGRLFIGDAHACQDGEI--CGTAVE 233
Db 275 MGGFIDMHVDLIK 287
QY 234 FASITTIKVLDLIK 246

RESULT 3
ID FTSZ_HALSA STANDARD; PRT; 375 AA.
AC Q48290;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE CELL DIVISION PROTEIN FTSZ HOMOLOG.
GN FTSZ.
OS Halobacterium salinarium.
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
OC Halobacterium.
RN [1]

```

```

SEQUENCE FROM N.A.
RP STRAIN=PHO81;
RX MEDLINE; 96200101.
RA Margolin W., Wang R., Kumar M.;
RT "Isolation of an ftsz homolog from the archaeobacterium Halobacterium
RT salinarium: implications for the evolution of FtsZ and tubulin.";
RL J. Bacteriol. 178:1320-1327(1996).
CC -1- FUNCTION: THIS PROTEIN IS ESSENTIAL TO THE CELL-DIVISION PROCESS.
CC ITS SEEMS TO ASSEMBLE INTO A DYNAMIC RING ON THE INNER SURFACE OF
CC THE CYTOPLASMIC MEMBRANE AT THE PLACE WHERE DIVISION WILL OCCUR,
CC AND THE FORMATION OF THE RING IS THE SIGNAL FOR SEPTATION TO
CC BEGIN. BINDS TO AND HYDROLYZES GTP (BY SIMILARITY).
CC -1- SUBUNIT: AGGREGATE TO FORM A RING-LIKE STRUCTURE (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE FTSZ FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U32860; AAB06191.1; -.
CC HSPP; Q57816; IFSZ.
CC PFAM; PF00091; tubulin; 1.
CC PRINTS; PR00423; CELLDIVISFTSZ.
CC PROSITE; PS01134; FTSZ_1; 1.
CC PROSITE; PS01135; FTSZ_2; 1.
CC Cell division; Septation; GTP-binding.
CC NP_BIND 105 113 GTP (POTENTIAL).
CC SQ SEQUENCE 375 AA; 39261 MW; 43A05013AB3E5BEB CRC64;

Query Match 5.2%; Score 125; DB 1; Length 375;
Best Local Similarity 23.8%; Pred. No. 1.33e-04;
Matches 34; Conservative 35; Mismatches 65; Indels 9; Gaps 8;

Db 186 DQITAEVKGISSETITQPSLINLDYADMTAIMNQGVAVLVGETQDKNKTNEVVKDAM 245
QY 51 DRIIVTDFAEAGINSQDIPSOQLKMPFLNPQNG-PIM-VNGA-EKGDVLAVIESML 107
Db 246 HPLLDVDFRGASGLVHTIG--GPDLTKEAGIADNI-TERLDASANVINGARTQESYK 302
QY 108 PRGVD-PY-GICAMTPHFGLTGTDLTAMLNDPLPEKVRMIKLD-SKVYVNSKRHTLYK 164
Db 303 GKVRVMAIMTGVQSAQVLGPSTQ 325
QY 165 PHIGTSLVSPSEIDSINSLTPDNH 187

RESULT 4
ID MEFA_MOUSE STANDARD; PRT; 498 AA.
AC Q60929;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE MYOCYTE-SPECIFIC ENHANCER FACTOR 2A.
GN MEF2A.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=CEREBELLUM;
RX MEDLINE; 97165895.
RA Lin X., Shah S., Bulleit R.F.;
RT "The expression of MEF2 genes is implicated in CNS neuronal
RT differentiation.";
RN Brain Res. Mol. Brain Res. 42:307-316(1996).
CC -1- FUNCTION: TRANSCRIPTION FACTOR WHICH BINDS SPECIFICALLY TO THE
CC MEF2 ELEMENT PRESENT IN THE REGULATORY REGIONS OF MANY MUSCLE-
CC SPECIFIC GENES. ACTIVATES TRANSCRIPTION VIA THIS ELEMENT. MAY BE
CC INVOLVED IN MUSCLE-SPECIFIC AND/OR GROWTH FACTOR-RELATED

```



```
CC CC TRANSCRIPTION (BY SIMILARITY).
CC CC -!- SUBUNIT: HETERODIMER.
CC CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC CC -!- SIMILARITY: BELONGS TO THE MADS DOMAIN FAMILY OF TRANSCRIPTION
CC CC FACTORS. MEF2 SUBFAMILY.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL: U30823; AAA74030.1; -.
CC CC DR HSSP; P11831; 1SR5.
CC CC DR MGD; MGI:99532; MEF2A.
CC CC DR PFAM; PF00319; SRF-TF; 1.
CC CC DR PRINTS; PR00404; MADSDOMAIN.
CC CC DR PROSITE; PS00350; MADS_BOX_1; 1.
CC CC DR PROSITE; PS00666; MADS_BOX_2; 1.
CC CC KW Transcription regulation; Nuclear protein; DNA-binding; Activator;
CC CC Multigene family.
CC CC FT DOMAIN 3 57 MADS.
CC CC FT DNA_BIND 58 86 MEF2-TYPE (POTENTIAL).
CC CC FT DOMAIN 254 257 POLY-PRO.
CC CC FT DOMAIN 288 293 POLY-GLU.
CC CC FT DOMAIN 419 423 POLY-HIS.
CC CC FT DOMAIN 448 455 POLY-SER.
CC CC SQ SEQUENCE 498 AA; 53724 MW; 590678D1BD1B3723 CRC64;
Query Match 4.6%; Score 110; DB 1; Length 498;
Best Local Similarity 29.2%; Pred. No. 2.45e-02;
Matches 19; Conservative 20; Mismatches 25; Indels 1; Gaps 1;
Db 107 HSPLED-REIKLNEUSDFTFKRGPCFPQNFMSVTVTPNLSLDTNPGSSLSVSPS 165
QY 136 NDPLEKVRMIKLDSEKVVYMKRHTLPYKPHIGTSLVSPEIDSLNTPDNGHGMVDVP 195
Db 166 LAASS 170
QY 196 IGPSS 200
RESULT 5
ID YF46_AQUAE STANDARD; PRT; 322 AA.
AC O67500;
DT 15-FEB-2000 (Rel. 39, Created)
DT 15-FEB-2000 (Rel. 39, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE HYPOTHETICAL PROTEIN AQ_1546.
GN AQ_1546.
OS Aquifex aeolicus.
OC Bacteria; Aquificales; Aquificaceae; Aquifex.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE; 98196666.
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus."
RL Nature 392:353-358(1998).
CC -!- SIMILARITY: BELONGS TO THE SIS FAMILY. GUTQ/KPSF SUBFAMILY.
CC CC -!- SIMILARITY: CONTAINS 2 CBS DOMAINS.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
```

```
CC CC or send an email to license@isb-sib.ch).
CC CC EMBL: AE000745; AAC07460.1; -.
CC CC DR PFAM; PF00571; CBS; 2.
CC CC DR PFAM; PF01380; SIS; 1.
CC CC KW Hypothetical protein; ATP-binding; Repeat; CBS domain.
CC CC FT NP_BIND 50 55 ATP (POTENTIAL).
CC CC FT DOMAIN 203 256 CBS 1.
CC CC FT DOMAIN 268 321 CBS 2.
CC CC SQ SEQUENCE 322 AA; 35114 MW; 74BE86D9B8FB2A69 CRC64;
Query Match 4.3%; Score 103; DB 1; Length 322;
Best Local Similarity 25.2%; Pred. No. 2.40e-01;
Matches 28; Conservative 35; Mismatches 41; Indels 7; Gaps 7;
Db 83 GDMGLDSKDALI-AISNSGESTVLVYQAKALNPVIGITGNEKSS-LAKYSDVVLK 140
QY 50 GDRIIVTRDAFEGAINSEQDIPSQLKMPFLNPQNGPIM-VNGAEKGDVLAVIESMLP 108
Db 141 IPVDR-EACP-F-NLAPTVSTVTALGDAITAMLMKLGFSQEDF-AKRH 187
QY 109 RGVDPYICAMIPHFGGLTGDTLAMLNDPLPEKVRMIKLDSEKVVWSKRH 159
RESULT 6
ID BTUC_ECOLI STANDARD; PRT; 326 AA.
AC P06609; P7197;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE VITAMIN B12 TRANSPORT SYSTEM PERMEASE PROTEIN BTUC.
GN BTUC.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 86304184.
RA Friedrich M.J., Deveaux L.C., Kadner R.J.;
RT "Nucleotide sequence of the btuCD genes involved in vitamin B12
RT transport in Escherichia coli and homology with components of
RT periplasmic-binding-protein-dependent transport systems."
RL J. Bacteriol. 167:928-934(1986).
RN [2]
RP REVISIONS.
RX MEDLINE; 89364713.
RA Rioux C.R., Kadner R.J.;
RT "Vitamin B12 transport in Escherichia coli K12 does not require the
RT btuE gene of the btuCED operon."
RL Mol. Gen. Genet. 217:301-308(1989).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12/MG1655;
RX MEDLINE; 97426617.
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner F.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:12453-1474(1997).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RA Alba H., Baba T., Fujita K., Hayashi K., Honjo A., Horluchi T.,
RA Ikemoto K., Inada T., Isono S., Isono S., Kitagawa M., Kasai H.,
RA Kashimoto K., Kim S., Kimura S., Kitagawa M., Kitakawa M., Makino K.,
RA Masuda S., Miki T., Mizobuchi K., Mori H., Motomura K., Nakamura Y.,
RA Nashimoto H., Nishio Y., Oshima T., Saito N., Sampei G., Seki Y.,
RA Tagami H., Takemoto K., Wada C., Yamamoto Y., Yano M.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: PART OF THE BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM
CC FOR VITAMIN B12. PROBABLY RESPONSIBLE FOR THE TRANSLLOCATION OF THE
CC SUBSTRATE ACROSS THE MEMBRANE.
```


OS Oryctolagus cuniculus (Rabbit).
 QC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=HEART MUSCLE;

[illegible]

```
CC      TETRAHYDROXYCHALCONE (ALSO TERMED NARINGENIN-CHALCONE OR CHALCONE)
CC      WHICH CAN UNDER SPECIFIC CONDITIONS SPONTANEOUSLY ISOMERIZE INTO
CC      NARINGENIN.
CC      -I- CATALYTIC ACTIVITY: 3 MALONYL-COA + 4-COUMAROYL-COA = 4 COA +
CC      NARINGENIN-CHALCONE + 3 CO(2).
CC      -I- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF
CC      FLAVONOIDS, A LARGE CLASS OF SECONDARY PLANT METABOLITES, MANY
CC      OF WHICH ARE BRIGHTLY COLORED.
CC      -I- SIMILARITY: BELONGS TO THE CHALCONE/SITLBENE SYNTHASES FAMILY.
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/ or
CC      send an email to license@isb-sib.ch).
-----
CC      EMBL; U15947; AAC49030.1; .
DR      PFAM; PF00195; Chal_stil_synt; 1.
DR      PROSITE; PS00441; CHALCONE_SYNTH; 1.
DR      Flavonoid biosynthesis; Transferase; Acyltransferase;
KW      Multigene family.
FT      ACT_SITE 170..170 BY SIMILARITY.
SQ      SEQUENCE 396 AA; 43926 MW; 352C478C1AF892B6 CRC64;
-----
Query Match 4.1%; Score 98; DB 1; Length 396;
Best Local Similarity 23.7%; Pred. No. 1.14e+00;
Matches 18; Conservative 23; Mismatches 33; Indels 2; Gaps 2;

Db    3 TTVTYV-LTDWTSRRAKLEGDAKILAIGTATPAGSWDTQTYDPDYFRITNSQHLEHKEK 61
      ||| | | | | :||| :||| | | :| |:| ::| :| | :
Cc    238 TTIKVDLIKWLSPRMENAEMISGSRPLEADTRIAYRDLTIYWLVDFGFQWDAA 296.

Db    62 FRICKSKTKRKHLV 77
Cc    297 YMLLSQCCKVRLGNMV 312
Cc    YQ

RESULT 14
ID     CAPB_BACAN          STANDARD;             PRT;   397 AA.
AC     P19580;
DT     01-FEB-1991 (Rel. 17, Created)
DI     01-FEB-1991 (Rel. 17, Last sequence update)
DE     01-FEB-1994 (Rel. 28, Last annotation update)
DN     CAPB PROTEIN.
GC     CAPB.
OS     Bacillus anthracis.
OG     Plasmid pTE702.
OC     Bacteria; Firmicutes; Bacillus/Clostridium group;
OC     Bacillus/Staphylococcus group; Bacillus.
RN     [1]
RP     SEQUENCE FROM N.A.
RX     MEDLINE; 89123146.
RT     Makino S.I., Uchida I., Terakado N., Sasakawa C., Yoshikawa M.;
RT     "Molecular characterization and protein analysis of the cap region,
RT     which is essential for encapsulation in Bacillus anthracis.";
RJ     J. Bacteriol. 171:722-730(1989)
CC     -I- FUNCTION: ESSENTIAL FOR THE SYNTHESIS OF THE POLYGLUTAMATE
CC     CAPSULE OF BACILLUS ANTHRAXIS. MAY FORM A POLYGLUTAMYL SYNTHETASE
CC     COMPLEX TOGETHER WITH CAPA AND CAPC PROTEINS.
CC     -I- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED.
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/ or
CC      send an email to license@isb-sib.ch).
-----
DR      EMBL; M24150; AAA22286.1; .
DR      PIR; A30091; A30091.
```

W A R E L L (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Thu Aug 24 18:51:52 2000; MasPar time 26.80 Seconds
Tabular output not generated. 848.426 Million cell updates/sec

Title: >US-09-214-679-2
Description: (1-328) from US09214679.pep
Perfect Score: 2417
Sequence: 1 MKWLEESIMAKRGVGRKP.....GNMVDPKRYTGAMLNKLLV 328

Scoring table: PAM 150
Gap 11

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl12

1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phase 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 47.366; Variance 79.598; scale 0.595

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	266	11.0	285	2 Q9WXX3	ACETAMIDASE, PUTATIVE.	1.52e-32
2	197	8.2	377	1 Q9YBQ1	377AA LONG HYPOTHETICA	3.16e-18
3	163	6.7	389	1 O58755	389AA LONG HYPOTHETICA	1.19e-11
4	110	4.6	2091	3 P78616	FATTY ACID SYNTHASE, B	1.64e-02
5	109	4.5	789	5 Q22211	SIMILARITY TO CYTOCHRO	2.34e-02
6	103	4.3	322	2 O67500	POLYSTALIC ACID CAPSUL	1.85e-01
7	103	4.3	393	2 O53272	NIFS-LIKE PROTEIN.	1.85e-01
8	103	4.3	436	10 Q40118	DELTA-15 LINEOYL DESAT	1.85e-01
9	105	4.3	473	2 O51454	TRANSCRIPTIONAL ACTIVA	9.38e-02
10	104	4.3	597	5 O45218	ALKYL-DIHYDROXYACETONE	1.32e-01
11	104	4.3	1874	2 O33765	IGAL1 PROTEASE (EC 3.4.	1.32e-01
12	102	4.2	243	5 O17308	GABA RECEPTOR SUBUNIT	2.60e-01
13	102	4.2	567	1 O26456	SENSORY TRANSDUCTION H	2.60e-01
14	101	4.1	4968	6 Q29621	CARDIAC RYANODINE RECE	3.03e-01
15	99	4.1	281	2 O30796	MXAE.	7.03e-01
16	98	4.1	387	10 Q92PP7	OMEGA-3 FATTY ACID DES	9.75e-01
17	99	4.1	396	10 O04970	CHALCONE SYNTHASE (EC	7.03e-01
18	98	4.1	899	3 O59868	CA++TRANSPORTING ATPA	9.75e-01
19	99	4.1	1056	1 O22855	HYPOTHETICAL PROTEIN.	7.03e-01
20	100	4.1	1061	5 O96998	L1156.2 PROTEIN.	5.06e-01

21	99	4.1	4967	4 Q92736	RYANODINE RECEPTOR 2,	7.03e-01
22	99	4.1	5035	6 Q29104	RYANODINE RECEPTOR.	7.03e-01
23	99	4.1	5035	6 Q29105	RYANODINE RECEPTOR.	7.03e-01
24	96	4.0	178	10 Q92PJ2	ADP-RIBOSYLATION FACTO	1.86e+00
25	96	4.0	181	10 Q48920	ADP-RIBOSYLATION FACTO	1.86e+00
26	96	4.0	181	10 Q22413	ADP-RIBOSYLATION FACTO	1.86e+00
27	97	4.0	252	1 Q9YD32	252AA LONG HYPOTHETICA	1.35e+00
28	97	4.0	333	2 Q46974	ECOVIII RESTRICTION EN	1.35e+00
29	96	4.0	373	2 O69849	HYPOTHETICAL 39.3 KD P	1.86e+00
30	96	4.0	380	10 Q23802	PLASTID OMEGA-3 FATTY	1.86e+00
31	96	4.0	393	2 P96736	YMSC PROTEIN.	1.86e+00
32	96	4.0	393	5 O97121	ECYSTEROID-INDUCIBLE	1.86e+00
33	97	4.0	396	2 O33098	NIFS-LIKE PROTEIN.	1.35e+00
34	96	4.0	396	10 Q04967	CHALCONE SYNTHASE (EC	1.86e+00
35	96	4.0	396	10 Q04968	CHALCONE SYNTHASE (EC	1.86e+00
36	97	4.0	484	2 Q9X3W3	HYPOTHETICAL 50.3 KD P	1.35e+00
37	97	4.0	638	2 Q46453	CIPA PROTEIN (FRAGMENT	1.35e+00
38	96	4.0	798	2 O34481	YRRC PROTEIN.	1.86e+00
39	97	4.0	1301	2 Q06307	ALPHA-AMYLASE (EC 3.2.	1.35e+00
40	95	3.9	165	10 Q9XHS7	HYPOTHETICAL 18.5 KD P	2.56e+00
41	95	3.9	391	10 Q92PP7	OMEGA-3 FATTY ACID DES	2.56e+00
42	95	3.9	396	10 Q04966	CHALCONE SYNTHASE (EC	2.56e+00
43	95	3.9	396	10 Q04971	CHALCONE SYNTHASE (EC	2.56e+00
44	95	3.9	972	14 Q9YJVO	STRUCTURAL POLYPROTEIN	2.56e+00
45	95	3.9	1249	11 Q64560	TRIPTEPTIDYLPEPTIDASE I	2.56e+00

ALIGNMENTS

RESULT	1	PRELIMINARY;	PRT;	285 AA.
ID	Q9WXX3;			
AC	Q9WXX3;			
DT	01-NOV-1999 (Tremblrel. 12, Created)			
DT	01-NOV-1999 (Tremblrel. 12, Last sequence update)			
DT	01-NOV-1999 (Tremblrel. 12, Last annotation update)			
DE	ACETAMIDASE, PUTATIVE.			
GN	TM0119.			
OS	Thermotoga maritima.			
OC	Bacteria; Thermotogales; Thermotoga.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE; 99287316.			
RA	NELSON K.E., CLAYTON R.A., GILL S.R., GWINN M.L., DODSON R.J.,			
RA	HAFT D.H., HICKEY E.K., PETERSON J.D., NELSON W.C., KETCHUM K.A.,			
RA	MCDONALD L., UTTERBACK T.R., MALEK J.A., LINHER K.D., GARRETT M.M.,			
RA	STEWART A.M., COTTON M.D., PRATT M.S., PHILLIPS C.A., RICHARDSON D.,			
RA	HEIDELBERG J., SUTTON G.G., FLEISCHMANN R.D., WHITE O., SALZBERG S.L.,			
RA	SMITH H.O., VENTER J.C., FRASER C.M.;			
RT	"Evidence for lateral gene transfer between Archaea and bacteria from			
RL	genome sequence of Thermotoga maritima.";			
RL	Nature 399:323-329(1999).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	NELSON K.E., CLAYTON R.A., GILL S.R., GWINN M.L., DODSON R.J.,			
RA	HAFT D.H., HICKEY E.K., PETERSON J.D., NELSON W.C., KETCHUM K.A.,			
RA	MCDONALD L., UTTERBACK T.R., MALEK J.A., LINHER K.D., GARRETT M.M.,			
RA	STEWART A.M., COTTON M.D., PRATT M.S., PHILLIPS C.A., RICHARDSON D.,			
RA	HEIDELBERG J., SUTTON G.G., FLEISCHMANN R.D., WHITE O., SALZBERG S.L.,			
RA	SMITH H.O., VENTER J.C., FRASER C.M.;			
RL	Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; A8001698; AAD35213.1; -			
SO	SEQUENCE 285 AA; 31196 MW; DE555325 CRC32;			
Query Match 11.0%; Score 266; DB 2; Length 285;				
Best Local Similarity 30.8%; Pred. No. 1.52e-32;				
Matches 52; Conservative 40; Mismatches 72; Indels 5; Gaps 5;				
Db	117	IPHPMVGICVAPQCEYEPTGAHRHGGNMDKEI-TENVTVHLPVFGALIALGDVH	175	
Qy	161	LPYKPHIGTSLVSPEDSINSLTPDNHGGNMDVPDIPGSGITYPL-VRAGGRLFIGDAH	219	
Db	176	ATMGDGEVCVSACEVPAKVVEIDVSKSKE-RIKWPVFTNDAYIIVISLPIEALKEVTR	234	

```

QY 220 ACQGDGEICGTAVEFASITTIKVDLIKNNQSWPRMENAENIMSIGSARPLEDATEIAYR 279
Db 235 ETY-WFIQRKTTPTFDAYMLASLSDVGVISOLVNPAPKAKARIPKYIF 282
QY 280 DLIYWLVEDFGEQW-DAYMILLSQCGKVRGLGNMVDPKYTVGAMLNKNLL 327

RESULT 2
ID Q9YEQ1 PRELIMINARY; PRT; 377 AA.
AC Q9YEQ1;
DT 01-NOV-1999 (TremBLrel. 12, Created)
DT 01-NOV-1999 (TremBLrel. 12, Last sequence update)
DE 01-NOV-1999 (TremBLrel. 12, Last annotation update)
DE 377AA LONG HYPOTHETICAL ACETAMIDASE.
GN APE0528.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Aeropyrum.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K1.
RX MEDLINE; 99310339.
RA KAWARABAYASI Y., HINO Y., HORIKAWA H., YAMAZAKI S., HAIKAWA Y.,
RA JIN-NO K., TAKAHASHI M., SEKINE M., BABA S., ANKAI A., KOSUGI H.,
RA HOSOYAMA A., FUKUI S., NAGAI Y., NISHIJIMA K., NAKAZAWA H.,
RA TAKAMIYA M., MASUDA S., FUNAHASHI T., TANAKA T., KUDOH Y.,
RA YAMAZAKI J., KUSHIDA N., OGUCHI A., AOKI K., KUBOTA K., NAKAMURA Y.,
RA NOMURA N., SAKO Y., KIKUCHI H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1.";
RL DNA Res. 6:83-101(1999).
DR EMBL; AP000059; BAA79495.1; -
SQ SEQUENCE 377 AA; 41052 MW; 6FA6B2AE CRC32;

Query Match 8.2%; Score 197; DB 1; Length 377;
Best Local Similarity 34.2%; Pred. NO. 3.16e-18;
Matches 40; Conservative 27; Mismatches 47; Indels 3; Gaps 3;

Db 146 LGVGVAPSKELLKIKEREDLLKRGFVLPTPEGAVPREPVASEGLRTIPPRENGG 205
QY 131 LTAMLDLPEKVRMIKLDSEKVKYWSKRHTLPYKPHIGTSLVSP-EIDSINSLTPDNHGG 189
Db 206 NLDVHRHSPGSKIYFPVEGA-LFSVGDHAHQDGVCGTAIDMGATATLRFQVI 261
QY 190 NMDVPDIPGSIYPLVRAPGRLE-IGDAHACQDGEICGTAVEFASITTIKVDLI 245

RESULT 3
ID O58755 PRELIMINARY; PRT; 389 AA.
AC O58755;
DT 01-AUG-1998 (TremBLrel. 07, Created)
DT 01-AUG-1998 (TremBLrel. 07, Last sequence update)
DT 01-JAN-1999 (TremBLrel. 09, Last annotation update)
DE 389AA LONG HYPOTHETICAL PROTEIN.
GN PH1041.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-OT3;
RX MEDLINE; 98344137.
RA KAWARABAYASI Y., SAWADA M., HORIKAWA H., HAIKAWA Y., HINO Y.,
RA YAMAMOTO S., SEKINE M., BABA S., KOSUGI H., HOSOYAMA A., NAGAI Y.,
RA SAKAI M., OGURA K., OTUKA R., NAKAZAWA H., TAKAMIYA M., OHFUKU Y.,
RA FUNAHASHI T., TANAKA T., KUDOH Y., YAMAZAKI J., KUSHIDA N., OGUCHI A.,
RA AOKI K., NAKAMURA Y., ROBB T.F., HORIKOSHI K., MASUCHI Y., SHIZUYA H.,
RA KIKUCHI H.;
RT "Complete sequence and gene organization of the genome of a hyper-
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA Res. 5:55-76(1998).
DR EMBL; AP000004; BAA30139.1; -
SQ SEQUENCE 389 AA; 42005 MW; B4D3990F CRC32;

Query Match 6.7%; Score 163; DB 1; Length 389;

```

```

Best Local Similarity 38.7%; Pred. NO. 1.19e-11;
Matches 24; Conservative 17; Mismatches 21; Indels 0; Gaps 0;

Db 206 GHLDVDSVREGAIVAPKVEGGVYAGDAHAMEGDGEVAGHTTDTAETTVTVVEIKGL 265
QY 189 GNMVDPDIPGSIYPLVRAPGRLE-IGDAHACQDGEICGTAVEFASITTIKVDLIKNN 248

Db 266 NL 267
QY 249 QL 250

RESULT 4
ID P78616 PRELIMINARY; PRT; 2091 AA.
AC P78616;
DT 01-MAY-1997 (TremBLrel. 03, Created)
DT 01-MAY-1997 (TremBLrel. 03, Last sequence update)
DT 01-NOV-1999 (TremBLrel. 12, Last annotation update)
DE FATTY ACID SYNTHASE, BETA SUBUNIT.
GN FASB.
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Euscomycetes; Plecomycetes;
OC Eurotiales; Trichocomaceae; Emericella.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97121482.
RA BROWN D.W., ADAMS T.H., KELLER N.P.;
RT "Aspergillus has distinct fatty acid synthases for primary and
RT secondary metabolism.";
RL Proc. Natl. Acad. Sci. U.S.A. 93:14873-14877(1996).
DR EMBL; U75347; AAB41494.1; -
DR PFAM; PF00698; Acyl_transf; 1.
DR PFAM; PF01575; MaoC_like; 1.
SQ SEQUENCE 2091 AA; 232283 MW; 6B5D181B CRC32;

Query Match 4.6%; Score 110; DB 3; Length 2091;
Best Local Similarity 33.7%; Pred. NO. 1.64e-02;
Matches 28; Conservative 15; Mismatches 35; Indels 5; Gaps 4;

Db 467 IPVYGTGKDDLRAISDANVVPALVRMTIDHP--VNWQTTAFPNATHIVDFG-PGGISG 523
QY 120 IPHFGGLTGDTLAMLN-DLPEKVRMIKLDSEKVKYWSKRHTLPYKPHIGTSLVSP 178

Db 524 LGVLTNRNKDGT-GVRVILAGSM 545
QY 179 INSLTPDNHGMVDPDIPGSI 201

RESULT 5
ID Q22211 PRELIMINARY; PRT; 789 AA.
AC Q22211;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TremBLrel. 08, Last annotation update)
DE SIMILARITY TO CITOCROME B.
GN T05B11.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., COOPER J., COULSON A.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SWALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;

```



```
RL Plant Physiol. 108:861-861(1995).
DR EMBL; U17063; AAA86690.1; -.
DR MENDEL; 8699; Limdo:1208;8699.
DR PFAM; PF00487; FA_desaturase; 1.
SQ SEQUENCE 436 AA; 50093 MW; 327D77FE CRC32;

Query Match 4.3%; Score 103; DB 10; Length 436;
Best Local Similarity 22.9%; Pred. No. 1.85e-01;
Matches 24; Conservative 29; Mismatches 42; Indels 10; Gaps 9;

Db 83 GSPPPP-KLADIRAAIPKHCWKNQWRSYVVRDVIIVGLAAVAANSW-AWPL-Y 139
QY 229 GRAVEFASITTKVDLIKN-W-QLSPRMENAE-NIMSTI-G-SARPLEDATRIAYRDLIY 283
Db 140 WVAQGMF--WALFVIGHDCGHSFNHKLNSVVGHLHLSILV 182
QY 284 WLVEDFGFEQWDAYMLSCGKVRGLGNMVDPKYTVGAMLNKLLV 328

RESULT 9
ID Q51454 PRELIMINARY; PRT; 473 AA.
AC Q51454;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE TRANSCRIPTIONAL ACTIVATOR.
GN FLER.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
OC Pseudomonas.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PAK;
RX MEDLINE; 96071911.
RA RITCHINGS B.W., ALMIRA E.C., LORY S., RAMPHAL R.;
RT "Cloning and phenotypic characterization of fliC and fliR, new
RT response regulators of Pseudomonas aeruginosa which regulate motility
RT and adhesion to mucin."
RL Infect. Immun. 63:4868-4876(1995).
DR EMBL; L41213; AAA93530.1; -.
DR HSSP; L41789; INTR.
DR PROSITE; PS00675; SIGMA54_INTERACT_1; 1.
DR PROSITE; PS00676; SIGMA54_INTERACT_2; 1.
DR PROSITE; PS00688; SIGMA54_INTERACT_3; 1.
DR PFAM; PF00072; response_reg; 1.
DR PFAM; PF00158; sigma54; 1.
SQ SEQUENCE 473 AA; 51264 MW; BA045A5 CRC32;

Query Match 4.3%; Score 105; DB 2; Length 473;
Best Local Similarity 30.8%; Pred. No. 9.38e-02;
Matches 20; Conservative 18; Mismatches 21; Indels 6; Gaps 6;

Db 128 DGPVALEPA-SRQLELAARVARDSTVLISGSGTGKEVLANYTHQOSPRAKPFITAIN 186
QY 62 EGAINSEQDIFSLLKMPF-LNPQNGPIWNG-AEKG-DVLAVIESMLPRGVDPY-GI- 116
Db 187 CAAP 191
QY 117 CAMIP 121

RESULT 10
ID O45218 PRELIMINARY; PRT; 597 AA.
AC O45218;
DT 01-JUN-1998 (Tremblrel. 06, Created)
DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE ALKYL-DIHYDROXYACETONEPHOSPHATE SYNTHASE (EC 2.5.1.26)
DE (ALKYLGLYCERONE-PHOSPHATE SYNTHASE)
DE (ALKYLDIHYDROXYACETONEPHOSPHATE SYNTHASE).
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.

RL SEQUENCE FROM N.A.
RP STRAIN=CB1489 HIM-8(E1489); TISSUE=WHOLE ANIMAL;
RX MEDLINE; 98113342.
RA DE VET E.C.J.M., PRINSEN H.C.M.T., VAN DEN BOSCH H.;
RT "Nucleotide sequence of a cDNA clone encoding a Caenorhabditis elegans
RT homolog of mammalian alkyl-dihydroxyacetonephosphate synthase:
RT evolutionary switching of peroxisomal targeting signals."
RL Biochem. Biophys. Res. Commun. 242:277-281(1998).
CC -1- CATALYTIC ACTIVITY: 1-ACYL-GLYCERONE 3-PHOSPHATE + A LONG-CHAIN
CC ALCOHOL = 1-ALKYL-GLYCERONE 3-PHOSPHATE + A LONG-CHAIN ACID ANION.
DR EMBL; AJ002686; CAA05690.1; -.
DR PFAM; PF01565; FAD_binding_4; 1.
KW Transference.
SQ SEQUENCE 597 AA; 66559 MW; AFB98BEE CRC32;

Query Match 4.3%; Score 104; DB 5; Length 597;
Best Local Similarity 28.2%; Pred. No. 1.32e-01;
Matches 24; Conservative 25; Mismatches 29; Indels 7; Gaps 5;

Db 159 CAIIP-IGG--GTSVTNALDTPETEKRAVISMDALLDKILWIDRENLTCAQAQIVGQS 215
QY 117 CAMIPFGGLTGTDLTAMLNPLPEKVRMIKLD-S--EKVYWSKRHTLPKPHIGTILSVS 173
Db 216 LERQLNKKGFTCGHEPDSIEFSTLG 240
QY 174 PEID-SINSLTPDNHGGNMVDYDIG 197

RESULT 11
ID O33765 PRELIMINARY; PRT; 1874 AA.
AC O33765;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE IGAI PROTEASE (EC 3.4.24.13).
GN IGA.
OS Streptococcus sanguis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SR85;
RX MEDLINE; 98084472.
RA POULSEN K., REINHOLDT J., JESPERGAARD C., BOYE K., BROWN T.A.,
RA HAUGE M., KILIAN M.;
RT "A comprehensive genetic study of streptococcal immunoglobulin A1
RT proteases: evidence for recombination within and between species."
RT Infect. Immun. 66:181-190(1998).
DR EMBL; Y13461; CAA73858.1; -.
DR PFAM; PF00746; Gram_pos_anchor; 1.
KW Protease; Hydrolase; Metalloprotease.
SQ SEQUENCE 1874 AA; 208478 MW; 4EF17F74 CRC32;

Query Match 4.3%; Score 104; DB 2; Length 1874;
Best Local Similarity 29.4%; Pred. No. 1.32e-01;
Matches 25; Conservative 21; Mismatches 33; Indels 6; Gaps 6;

Db 399 GGIOGAIVEPEQVSSLPPEYTCPOAGAV-VE-PEQVAPLAEYIGPQAGSVVEPQVPL- 455
QY 63 GAINSEQDI-PSQLLKMP-FLNPQNGPIWNGAEKGDVLAIVIESMLPRGVDPYGCAMI 120
Db 456 PEYTGVOAGSVWSPEQATPLPEYTR 480
QY 121 PHFGGL-TGTDLTAMLNPLPEKVR 144

RESULT 12
ID O17308 PRELIMINARY; PRT; 243 AA.
AC O17308;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-NOV-1999 (Tremblrel. 08, Last annotation update)
```

```

RESULT      14
ID          Q29621 PRELIMINARY; PRT; 4968 AA.
AC          Q29621;
DT          01-NOV-1996 (TREMBLrel. 01, Created)
DT          01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT          01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE          CARDIAC RYANODINE RECEPTOR.
GN          RYR-2.
OS          Oryctolagus cuniculus (Rabbit).
OC          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC          Eutheria; Lagomorpha; Leporidae; Oryctologus.
[1]
RN          SEQUENCE FROM N.A.
RC          STRAIN-DOMESTIC RABBIT; TISSUE=HEART, BRAIN;
RX          MEDLINE; 91032060.
RA          NAKAI J., IMAGAWA T., HAKAMAT Y., SHIGEKAWA M., TAKESHIMA H., NUWA S.;
RT          "Primary structure and functional expression from cDNA of the cardiac
RT          ryanodine receptor/calcium release channel.";
RL          FEBS Lett. 271:169-177(1990).
RN          [2]
RS          SEQUENCE FROM N.A.
RC          TISSUE=HEART, BRAIN;
RX          MEDLINE; 93147068.
RA          IMAGAWA T., NAKAI J., TAKESHIMA H., NAKASAKI Y., SHIGEKAWA M.;
RT          "Expression of Ca(2+)-induced Ca2+ release channel activity from
RT          cardiac ryanodine receptor cDNA in Chinese hamster ovary cells.";
RL          J. Biochem. 112:508-513(1992).
RN          [3]
RS          SEQUENCE FROM N.A.
RC          STRAIN-DOMESTIC RABBIT; TISSUE=HEART, BRAIN;
RA          NAKAI J.;
RT          Submitted (FEB-1996) to the EMBL/GenBank/DDBJ databases.
DL          EMBL; U50465; AAA93465.1; -.
DR          DDB; U50466; AAA93469.1; -.
DR          PFAM; PF01365; RYDR_ITPR; 1.
DR          PFAM; PF00822; SPRY; 3.
DR          PRINTS; PR00795; RYANODINER.
SQ          SEQUENCE 4968 AA; 565081 MW; FDEFSE60 CRC32;
Query Match 4.2%; Score 101; DB 6; Length 4968;
Best Local Similarity 26.6%; Pred.No. 3 63e-01;
Matches 17; Conservative 19; Mismatches 24; Indels 4; Gaps

Db          3385 LKEPTPEAEELFRMAEVFIYWKSQSHN--FKREEQNFYVQNEINNMSFLITDTKS-KMWSK 3441
              ::::| |::::| :||| ||:| :| :| ::::| | :| :|
OY          135 LNPLPEKVRYMKLDSEK-VYWSKRHTLPYKHICTLSVSPEIDSINSLSLTDPNHGGMNV 193
              :::
Db          3442 AAVS 3445
OY          194 PDIG 197

RESULT      15
ID          Q30796 PRELIMINARY; PRT; 281 AA.
AC          Q30796;
DT          01-JAN-1998 (TREMBLrel. 05, Created)
DT          01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT          01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE          MXAE.
DX          MXAE.
GN          MXAE..
OS          Methylobacterium extorquens.
OC          Bacteria; Proteobacteria; alpha subdivision; Methylobacterium.
[1]
RN          SEQUENCE FROM N.A.
RC          STRAIN=AM1.
RX          MEDLINE; 98156126.
RA          SPRINGER A.L., AUMAN A.J., LIDSTROM M.E.;
RT          "Sequence and characterization of mxaB, a response regulator involved
RT          in regulation of methanol oxidation, and of mxaW, a methanol-regulated
RT          gene in Methylobacterium extorquens AM1.";
RL          FEMS Microbiol. Lett. 160:119-124(1998).
DR          EMBL; AF017434; AAC46163.1; -.
SQ          SEQUENCE 281 AA; 28758 MW; 5C08AFFA CRC32;
```

Query Match 4.1%; Score 99; DB 2; Length 281;
Best Local Similarity 25.9%; Pred. No. 7.03e-01;
Matches 21; Conservative 23; Mismatches 33; Indels 4; Gaps 4;
Db 127 DGAWRTRLATIPVG-TAPFALALSPDERRLYVG-NVRS-NDLTVIDTGTGLKATATVPAGA 183
QY 141 EKVRMIKLDSEKVVWSRRHTLPYKPHIGTSLVSPEIDSLNSLTPDNHGGNMDVPDIPGGS 200
Db 184 MPYGVSVSPDGDARVFTNQHA 204
QY 201 ITVPL-VRAPGGRLEICDAHA 220

Search completed: Thu Aug 24 18:53:01 2000
Job time : 69 secs.